L Number	Hits	Search Text	DB	Time stamp
1	1188	exon$1 same sequenc$ same align$	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/05/12 18:59
2	1385	702/19[ccls]	USPAT; US-PGPUB	2004/05/12 18:59
3	898	702/20[ccls]	USPAT; US-PGPUB	2004/05/12 18:59
4	2107	702/19[ccls] or 702/20[ccls]	USPAT; US-PGPUB	2004/05/12 18:59
5	29	(exon$1 same sequenc$ same align$) and (702/19[ccls] or 702/20[ccls])	USPAT; US-PGPUB	2004/05/12 19:01
6	9	exon$ and sequenc$ and align$	EPO; JPO; DERWENT; IBM_TDB	2004/05/12 19:02

********** Welcome to STN International ********
**

NEWS 1 Web Page URLs for STN Seminar Schedule - N. America
NEWS 2 "Ask CAS" for self-help around the clock
NEWS 3 JAN 27 Source of Registration (SR) information in REGISTRY updated
and searchable
NEWS 4 JAN 27 A new search aid, the Company Name Thesaurus, available in
CA/CAplus
NEWS 5 FEB 05 German (DE) application and patent publication number format
changes
NEWS 6 MAR 03 MEDLINE and LMEDLINE reloaded
NEWS 7 MAR 03 MEDLINE file segment of TOXCENTER reloaded
NEWS 8 MAR 03 FRANCEPAT now available on STN
NEWS 9 MAR 29 Pharmaceutical Substances (PS) now available on STN
NEWS 10 MAR 29 WPIFV now available on STN
NEWS 11 MAR 29 New monthly current-awareness alert (SDI) frequency in RAPRA
NEWS 12 APR 26 PROMT: New display field available
NEWS 13 APR 26 IFIPAT/IFIUDB/IFICDB: New super search and display field
available
NEWS 14 APR 26 LITALERT now available on STN
NEWS 15 APR 27 NLDB: New search and display fields available
NEWS 16 May 10 PROUSDDR now available on STN
NEWS 17 May 19 PROUSDDR: One FREE connect hour, per account, in both May
and June 2004
NEWS 18 May 12 EXTEND option available in structure searching
NEWS 19 May 12 Polymer links for the POLYLINK command completed in REGISTRY

NEWS EXPRESS MARCH 31 CURRENT WINDOWS VERSION IS V7.00A, CURRENT
MACINTOSH VERSION IS V6.0c(ENG) AND V6.0Jc(JP),
AND CURRENT DISCOVER FILE IS DATED 26 APRIL 2004
NEWS HOURS STN Operating Hours Plus Help Desk Availability
NEWS INTER General Internet Information
NEWS LOGIN Welcome Banner and News Items
NEWS PHONE Direct Dial and Telecommunication Network Access to STN
NEWS WWW CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that
specific topic.

All use of STN is subject to the provisions of the STN Customer agreement. Please note that this agreement limits use to scientific research. Use for software development or design or implementation
of commercial gateways or other similar uses is prohibited and may result in loss of user privileges and other penalties.

********** STN Columbus **********
**

FILE 'HOME' ENTERED AT 18:51:37 ON 12 MAY 2004

=> file caplus
COST IN U.S. DOLLARS SINCE FILE ENTRY TOTAL SESSION
FULL ESTIMATED COST 0.21 0.21

FILE 'CAPLUS' ENTERED AT 18:51:46 ON 12 MAY 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

FILE COVERS 1907 - 12 May 2004 VOL 140 ISS 20
FILE LAST UPDATED: 11 May 2004 (20040511/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

=> s (exon#(10a)sequenc?(10a)align?)/bi,ab
45895 EXON#/BI
42817 EXON#/AB
726837 SEQUENC?/BI
613132 SEQUENC?/AB
96740 ALIGN?/BI
87143 ALIGN?/AB
L1 177 (EXON#(10A)SEQUENC?(10A)ALIGN?)/BI,AB

=> s l1 and algorithm?/bi,ab
76518 ALGORITHM?/BI
65008 ALGORITHM?/AB
L2 6 L1 AND ALGORITHM?/BI,AB

=> s l1 not 2004/py
385703 2004/PY
L3 121 L1 NOT 2004/PY

=> s l3 not 2003/py
1135324 2003/PY
L4 110 L3 NOT 2003/PY

=> s l4 not 2002/py
1127076 2002/PY
L5 102 L4 NOT 2002/PY

=> s l5 not 2001/py
1076644 2001/PY
L6 68 L5 NOT 2001/PY

=> d l6 1-68 bib ab

L6 ANSWER 1 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:871401 CAPLUS Full-text
DN 135:56820
TI Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of
homotypic trans-splicing
AU Takahara, Terunao; Kanazu, Shin-Ichi; Yanagisawa, Shuichi; Akanuma,
Hiroshi
CS Department of Life Sciences (Chemistry), Graduate School of Arts and
Sciences, The University of Tokyo, Tokyo, 153-8902, Japan
SO Journal of Biological Chemistry (2000), 275(48), 38067-38072

CODEN: JBCHA3; ISSN: 0021-9258
PB American Society for Biochemistry and Molecular Biology
DT Journal
LA English
AB Sp1 is one of the well documented transcription factors, but the whole structure of human Sp1 has not been determined yet. In the present study, the authors isolated several cDNAs representing two forms of human Sp1 mRNA with different 5'-terminal structures in HepG2 cells. Isolation of a genomic clone established that one of the cDNAs represents the mRNA having consecutive **alignment** of **exons**, which allowed deducing the complete amino acid **sequence** for human Sp1. Another cDNA clone had a surprising structure that possessed an alignment of exons 3-2-3. Both reverse transcriptase-polymerase chain reaction and RNase protection assays confirmed accumulation of the two forms of Sp1 mRNA in HepG2 cells. Because Southern blot anal. suggested that exon 3 exists as a single copy in the genome, the cDNA clone having the duplicated sequences for exon 3 appeared to reflect the trans-splicing between pre-mRNAs of human Sp1.
RE.CNT 38 THERE ARE 38 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 2 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:820417 CAPLUS Full-text
DN 134:348690
TI Sequence-based typing of HLA-B: The B7 cross-reacting group
AU Voorter, C. E. M.; van der Vlies, S. A.; van den Berg-Loonen, E. M.
CS Tissue Typing Laboratory, University Hospital Maastricht, Maastricht, 6202
AZ, Neth.
SO Tissue Antigens (2000), 56(4), 356-362
CODEN: TSANA2; ISSN: 0001-2815
PB Munksgaard International Publishers Ltd.
DT Journal
LA English
AB The large no. of polymorphic sites in the HLA-B locus makes sequencing an efficient way of detecting and analyzing them. Most polymorphic sites are located in the α1 and α2 domains of the mol., encoded by exons 2 and 3 of the gene. An HLA-B-specific sequence-based typing (SBT) strategy was designed for routine application identifying the polymorphic sites in these domains. Exons 2 and 3 were amplified sep. using amplification primers located in intron 1, intron 2 and intron 3. Sep. amplification of exons 2 and 3 resulted in short polymerase chain reacting (PCR) products and enabled a solid-phase sequencing approach, which made correct assignment of heterozygous positions possible due to low background. A one-step sequencing reaction was performed using fluorescent dye-labeled sequencing primers. One forward sequencing reaction was performed for exon 2, whereas for exon 3, two forward sequencing reactions were needed using two different sequencing primers located in intron 2 and exon 3. The combined **sequences** of **exon** 2 and 3 were used for automatic **alignment** to an HLA-B sequence database and automatic allele assignment. A total of 355 individuals with at least one allele belonging to the B7 cross-reacting group (B7, 13, 22, 27, 40, 41, 42, 47, 48, 81 and 82) were typed for HLA-B by SBT. In the B7 group 48 different alleles were identified, in the non-B7 group a further 59 alleles were sequenced, 9 new alleles were identified. The sequencing strategy described has proven to be reliable and efficient for high-resolution HLA-B typing.
RE.CNT 23 THERE ARE 23 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 3 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:736627 CAPLUS Full-text
DN 134:290985
TI Amino acid and nucleotide recurrence in aligned sequences: synonymous
substitution patterns in association with global and local base compositions
AU Nishizawa, Manami; Nishizawa, Kazuhisa
CS Department of Biochemistry, Teikyo University School of Medicine, Kaga,
Itabashi, Tokyo, 173, Japan
SO Nucleic Acids Research (2000), 28(19), 3801-3810
CODEN: NARHAD; ISSN: 0305-1048
PB Oxford University Press
DT Journal
LA English
AB The tendency for repetitiveness of nucleotides in DNA sequences has been reported for a variety of organisms. We show that the tendency for repetitive use of amino acids is widespread and is observed even for segments conserved between human and Drosophila melanogaster at the level of >50% amino acid identity. This indicates that repetitiveness influences not only the weakly constrained segments but also those sequence segments conserved among phyla. Not only glutamine (Q) but also many of the 20 amino acids show a comparable level of repetitiveness. Repetitiveness in bases at codon position 3 is stronger for human than for D. melanogaster, whereas local repetitiveness in intron sequences is similar between the two organisms. While genes for immune system-specific proteins, but not ancient human genes (i.e. human homologs of Escherichia coli genes), have repetitiveness at codon bases 1 and 2, repetitiveness at codon base 3 for these groups is similar, suggesting that the human genome has at least two mechanisms generating local repetitiveness. Neither amino acid nor nucleotide repetitiveness is observed beyond the exon boundary, denying the possibility that such repetitiveness could mainly stem from natural selection on mRNA or protein sequences. Analyses of mammalian sequence alignments show that while the 'between gene' GC content heterogeneity, which is linked to 'isochores', is a principal factor associated with the bias in substitution patterns in human, 'within gene' heterogeneity in nucleotide composition is also associated with such bias on a more local scale. The relationship amongst the various types of repetitiveness is discussed.
RE.CNT 45 THERE ARE 45 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 4 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:455041 CAPLUS Full-text
DN 134:54626
TI Molecular and Cytogenetic Analysis of Lymphoblastoid and Colon Cancer Cell
Lines From Cotton-top Tamarin (Sagiunus oedipus)
AU Mao, X.; McGuire, S.; Hamoudi, R. A.
CS Human Cytogenetics Laboratory, Imperial Cancer Research Fund, London, UK
SO Cancer Genetics and Cytogenetics (2000), 120(1), 6-10
CODEN: CGCYDF; ISSN: 0165-4608
PB Elsevier Science Inc.
DT Journal
LA English
AB The cotton-top tamarin (CTT) (Sagiunus oedipus) has been used as an animal model to investigate the etiol. and pathophysiol. of several human diseases, including

ulcerative colitis and its associated colorectal carcinoma (CRC). Little is known, however, about genetic synteny between CTT and humans, and about chromosome aberrations in CTT CRC. To address these issues, we have analyzed CTT lymphoblastoid and CRC cell lines using cytogenetics, fluorescence in situ hybridization (Zoo-FISH), and direct sequencing. The CTT lymphocytes had pseudodiploid chromosomes of 46. The CTT CRC cells showed near-diploid chromosomes of 45. Several clonal structural aberrations were observed, including der(1), a marker chromosome, and double minutes. Zoo-FISH using human chromosome 2, 3, 5, 6, 9, 11, 13, 15, 16, 17, 19, 22, and X paints identified homologous chromosomes and subchromosomal regions in the CTT genome. Fluorescence in situ hybridization with human telomeric probe also detected a homologous sequence in CTT genome. Direct sequencing of CTT genomic DNA using primers amplifying exons 4 and 15 of the human APC gene identified DNA sequences in CTT genome with 99% and 95% homol., resp. These results provide a basis for further comparative studies of CTT and human genome.

RE.CNT 16 THERE ARE 16 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 5 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:306820 CAPLUS Full-text
DN 133:218242
TI Genie-Gene finding in Drosophila melanogaster
AU Reese, Martin G.; Kulp, David; Tammana, Hari; Haussler, David
CS Berkeley Drosophila Genome Project, Department of Molecular and Cell
Biology, University of California, Berkeley, CA, 94720-3200, USA
SO Genome Research (2000), 10(4), 529-538
CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB A hidden Markov model-based gene-finding system called Genie was applied to the genomic Adh region in Drosophila melanogaster as a part of the Genome Annotation Assessment Project (GASP). Predictions from three versions of the Genie gene-finding system were submitted, one based on statistical properties of coding genes, a second included EST alignment information, and a third that integrated protein sequence homol. information. All three programs were trained on the provided Drosophila training data. In addition, promoter assignments from an integrated neural network were submitted. The gene assignments overlapped >90% of the 222 annotated genes and 26 possibly novel genes were predicted, of which some might be overpredictions. The system correctly identified the exon boundaries of 70% of the exons in cDNA-confirmed genes and 77% of the **exons** with the addition of EST **sequence alignments** . The best of the three Genie submissions predicted 19 of the annotated 43 gene structures entirely correct (44%). In the promoter category, only 30% of the transcription start sites could be detected, but by integrating this program as a sensor into Genie the false-pos. rate could be dropped to 1/16,786 (0.006%). The results of the experiment on the long contiguous genomic sequence revealed some problems concerning gene assembly in Genie. The results were used to improve the system. The authors show that Genie is a robust hidden Markov model system that allows for a generalized integration of information from different sources such as signal sensors (splice sites, start codon, etc.), content sensors (**exons**, introns, intergenic) and **alignments** of mRNA, EST, and peptide **sequences**. The assessment showed that Genie could effectively be used for the annotation of complete genomes from higher organisms.

RE.CNT 11 THERE ARE 11 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 6 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:295356 CAPLUS Full-text
DN 133:203518
TI Prediction of the exon-intron structure by comparison of genomic sequences
AU Novichkov, P. S.; Gelfand, M. S.; Mironov, A. A.
CS State Research Center GosNIIGenetika, Moscow, 113545, Russia
SO Molecular Biology (Translation of Molekulyarnaya Biologiya (Moscow))
(2000), 34(2), 200-206
CODEN: MOLBBJ; ISSN: 0026-8933
PB MAIK Nauka/Interperiodica Publishing
DT Journal
LA English
AB An algorithm for prediction of the exon-intron structure of higher eukaryotic genes is suggested. The algorithm is based on comparison of genomic sequences of homologous genes from different species. It uses the fact that protein-coding sequences evolve slower than noncoding regions. Unlike the existing comparison methods, the proposed algorithm, which is a modified version of splicing alignment, compares not nucleotide but amino acid sequences, which increases its sensitivity. Conservation of the exon-intron structures of the compared genes is not assumed. The algorithm is implemented in the program Pro-Gen. The testing of the algorithm demonstrated that it can be successfully applied to prediction of vertebrate genes, and in some cases, for more distant comparisons (e.g., vertebrates and insects or nematodes). Thus, the program can be used for prediction of human genes by comparison with genes of model organisms: mouse, fugu, drosophila, and nematode. The algorithm overcomes deficiencies of the existing methods, both statistical (insufficient reliability) and similarity-based (inapplicability to completely new genes).

RE.CNT 13 THERE ARE 13 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 7 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:6192 CAPLUS Full-text
DN 132:303956
TI A novel approach for the computer analysis and allele assignment of
complex HLA class I sequences
AU Johnston-Dow, L.; Conrad, M.; Kronick, M.
CS Applied Biosystems Division of Perkin Elmer, Foster City, CA, 944404, USA
SO HLA: Genetic Diversity of HLA Functional and Medical Implication,
[Proceedings of the International Histocompatibility Workshop and
Conference], 12th, Saint-Malo and Paris, France, 1996 (1997), Meeting Date
1996, Volume 2, 365-366. Editor(s): Charron, Dominique. Publisher: EDK,
Medical and Scientific International Publisher, Sevres, Fr.
CODEN: 68MRA5
DT Conference
LA English
AB An approach was developed for anal. of sequencing-based typing data for HLA class I genes containing informative regions that span 2-3 exons which require 2-6 sep.

sequencing reactions. Factura HLA (Perkin Elmer) was used to screen each sequence against an exon-specific sequence to eliminate any intron region from subsequence consideration. In addition, the relative peak heights at each position were determined and International Union of Biol. ambiguity codes were assigned to heterozygous positions. The resulting sequences were assembled into a project using Sequence Navigator (Perkin Elmer) and a template including a consensus **sequence** for the gene of interest from **exons** 1-4 in a **Sequence** Navigator multiple **sequence alignment** layout. This assemblage is then aligned to the gene consensus for inspection and editing in Sequence Navigator.

RE.CNT 3 THERE ARE 3 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 8 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:6166 CAPLUS Full-text
DN 132:306827
TI Conserved sequence motifs create a pattern of MHC genetic diversification
within primate DRB lineages
AU Gaur, L. K.; Nepom, G. T.; Snyder, K. E.; Anderson, J.; Heise, E. R.
CS Molecular Biology laboratory, Puget Sound Blood Center, Seattle, WA,
98104-1256, USA
SO HLA: Genetic Diversity of HLA Functional and Medical Implication,
[Proceedings of the International Histocompatibility Workshop and
Conference], 12th, Saint-Malo and Paris, France, 1996 (1997), Meeting Date
1996, Volume 2, 274-276. Editor(s): Charron, Dominique. Publisher: EDK,
Medical and Scientific International Publisher, Sevres, Fr.
CODEN: 68MRA5
DT Conference; General Review
LA English
AB A review and discussion with 10 refs. Interspecies comparative studies among various nonhuman primates and humans are presented in order to analyze the generation and maintenance of specific localized MHC polymorphisms. Although several HVR1 (hypervariable region) sequences are conserved between human and nonhuman primates, consistent with the trans-species mode of inheritance, many other HVRI sequences are unique to the nonhuman primates. HVR **sequence alignments** from the second **exon** of human and other primate (4 species) DRB gene are presented. We propose that specific segmental interchange involving the HVRIII region has occurred among DR alleles, especially in the primate DRB6 lineage.

RE.CNT 10 THERE ARE 10 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 9 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1999:731730 CAPLUS Full-text
DN 132:235767
TI Characterisation of a novel HLA-A pseudogene, HLA-BEL, with significant
sequence identity with a gorilla MHC class I gene
AU Williams, F.; Curran, M. D.; Middleton, D.
CS Northern Ireland Regional Histocompatibility, City Hospital, Belfast, BT9
7TS, UK
SO Tissue Antigens (1999), 54(4), 360-369
CODEN: TSANA2; ISSN: 0001-2815
PB Munksgaard International Publishers Ltd.
DT Journal
LA English
AB During the development of an HLA-A polymerase chain reaction using sequence-specific oligonucleotide probes (PCR-SSOP) method for the identification of HLA-A*24 and -A*30 alleles, group amplification resulted in the formation of an unusual PCR product in certain individuals. This fragment was approx. 900 bp smaller than the expected product and was also detected in some non-HLA-A*24- and -A*30-pos. individuals acting as neg. controls for the group specific amplification. Nucleotide sequence anal. of this product identified it as a unique class I gene sequence displaying homol. to both primate and human class I A-locus genes. The entire gene was amplified using PCR and the complete DNA sequence information from exon 1 to exon 8, including introns, was determined A recombination event was identified which results in the fusion of intron 2 with intron 3, causing a deletion of the intervening exon 3 sequence. In addition, there are two cytosine insertions in the poly-cytosine stretch at the start of exon 4 which cause a frameshift and premature termination. The **exon** 1 and 2 **sequences** most closely **align** with the gorilla allele A*0501, displaying only five mismatches. PCR anal. has established that the gene is associated with the following HLA-A types: HLA-A*3001, -A*3301, -A*3303, -A*6802, -A*2901, -A*0203, -A*0205 and -A*31012. Reverse transcription (RT)-PCR anal. of individuals containing this gene failed to detect any mRNA transcription, suggesting that this is a previously undescribed non-expressed class I pseudogene which we have provisionally named HLA-BEL. Its unique gene structure gives a possible insight into the evolutionary pathway that created HLA class I genes.

RE.CNT 24 THERE ARE 24 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 10 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1999:581939 CAPLUS Full-text
DN 132:31433
TI Comparative Sequence Analysis of the Mouse and Human Lgn1/SMA Interval
AU Endrizzi, Matthew; Huang, Sidong; Scharf, Jeremiah M.; Kelter, Arndt-Rene;
Wirth, Brunhilde; Kunkel, Louis M.; Miller, Webb; Dietrich, William F.
CS Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA
SO Genomics (1999), 60(2), 137-151
CODEN: GNMCEP; ISSN: 0888-7543
PB Academic Press
DT Journal
LA English
AB Human chromosome 5q11.2-q13.3 and its ortholog on mouse chromosome 13 contain candidate genes for an inherited human neurodegenerative disorder called spinal muscular atrophy (SMA) and for an inherited mouse susceptibility to infection with Legionella pneumophila (Lgn1). These homologous genomic regions also have unusual repetitive organizations that create practical difficulties in mapping and raise interesting issues about the evolutionary origin of the repeats. In an attempt to analyze this region in detail, and as a way to identify addnl. candidate genes for these diseases, we have determined the sequence of 179 kb of the mouse Lgn1/SMA interval. We have analyzed this sequence using BLAST searches and various exon prediction programs to identify potential genes. Since these methods can generate false-pos. **exon** declarations, our **alignments** of the mouse **sequence** with available human orthologous sequence allowed us to

discriminate rapidly among this collection of potential coding regions by indicating which regions were well conserved and were more likely to represent actual coding sequence. As a result of our anal., we accurately mapped two addnl. genes in the SMA interval that can be tested for involvement in the pathogenesis of SMA. While no new Lgn1 candidates emerged, we have identified new genetic markers that exclude Smn as an Lgn1 candidate. In addition to providing important resources for studying SMA and Lgn1, our data provide further evidence of the value of sequencing the mouse genome as a means to help with the annotation of the human genomic sequence and vice versa. (c) 1999 Academic Press.

RE.CNT 68 THERE ARE 68 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 11 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1999:444895 CAPLUS Full-text
DN 131:197771
TI Characterization of the GAGE genes that are expressed in various human
cancers and in normal testis
AU De Backer, Olivier; Arden, Karen C.; Boretti, Mauro; Vantomme, Valerie; De
Smet, Charles; Czekay, Suzanne; Viars, Carrie S.; De Plaen, Etienne;
Brasseur, Francis; Chomez, Patrick; Van Den Eynde, Benoit; Boon, Thierry;
Van Der Bruggen, Pierre
CS Ludwig Institute for Cancer Research, Brussels Branch, and Cellular
Genetics Unit, Universite Catholique de Louvain, Brussels, B-1200, Belg.
SO Cancer Research (1999), 59(13), 3157-3165
CODEN: CNREA8; ISSN: 0008-5472
PB AACR Subscription Office
DT Journal
LA English
AB The GAGE-1 gene was identified previously as a gene that codes for an antigenic peptide, YRPRPRRY, which was presented on a human melanoma by HLA-Cw6 mols. and recognized by a clone of CTLs derived from the patient bearing the tumor. By screening a cDNA library from this melanoma, the authors identified five addnl., closely related genes named GAGE-2-6. The authors report here that further screening of this library led to the identification of two more genes, GAGE-7B and -8. GAGE-1, -2, and -8 code for peptide YRPRPRRY. Using another antitumor CTL clone isolated from the same melanoma patient, the authors identified antigenic peptide, YYWPRPRRY, which is encoded by GAGE-3, -4, -5, -6, and -7B and which is presented by HLA-A29 mols. Genomic cloning of GAGE-7B showed that it is composed of five exons. **Sequence alignment** showed that an addnl. **exon**, which is present only in the mRNA of GAGE-1, has been disrupted in gene GAGE-7B by the insertion of a long interspersed repeated element retroposon. These GAGE genes are located in the p11.2-p11.4 region of chromosome X. They are not expressed in normal tissues, except in testis, but a large proportion of tumors of various histol. origins express at least one of these genes. Treatment of normal and tumor cultured cells with a demethylating agent, azadeoxycytidine, resulted in the transcriptional activation of GAGE genes, suggesting that their expression in tumors results from a demethylation.

RE.CNT 52 THERE ARE 52 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 12 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1999:123645 CAPLUS Full-text
DN 130:321434
TI Comparison of Bombyx mori and Helicoverpa armigera cytoplasmic actin genes
provides clues to the evolution of actin genes in insects
AU Mange, Alain; Prudhomme, Jean-Claude
CS Centre de Genetique Moleculaire et Cellulaire, Universite Claude Bernard
Lyon I, Centre National de la Recherche Scientifique, Villeurbanne, F.
69622, Fr.
SO Molecular Biology and Evolution (1999), 16(2), 165-172
CODEN: MBEVEO; ISSN: 0737-4038
PB Society for Molecular Biology and Evolution
DT Journal
LA English
AB The cytoplasmic actin genes BmA3 and BmA4 of Bombyx mori were found clustered in a single genomic clone in the same orientation. As a similar clustering of the two cytoplasmic actin genes HaA3a and HaA3b also occurs in another lepidopteran, Helicoverpa armigera, we analyzed the sequence of the pair of genes from each species. Due to the high conservation of cytoplasmic actins, the coding sequence of the four genes was easily **aligned**, allowing the detection of similarities in noncoding **exon** and intron **sequences** as well as in flanking **sequences.** All four genes exhibited a conserved intron inserted in codon 117, an original position not encountered in other species. It can thus be postulated that all of these genes derived from a common ancestral gene carrying this intron after a single event of insertion. The comparison of the four genes revealed that the genes of B. mori and H. armigera are related in two different ways: the coding sequence and the intron that interrupts it are more similar between paralogous genes within each species than between orthologous genes of the two species. In contrast, the other (noncoding) regions exhibited the greatest similarity between a gene of one species and a gene of the other species, defining two pairs of orthologous genes, BmA3 and HaA3a on one hand and BmA4 and HaA3b on the other. However, in each species, the very high similarities of the coding sequence and of the single intron that interrupts it strongly suggest that gene conversion events have homogenized this part of the sequence. As the divergence of the B. mori genes was higher than that of the H. armigera genes, we postulated that the gene conversion occurred earlier in the B. mori lineage. This leads us to hypothesize that gene conversion could also be responsible for the original transfer of the common intron to the second gene copy before the divergence of the B. mori and H. armigera lineages.

RE.CNT 22 THERE ARE 22 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 13 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1999:28363 CAPLUS Full-text
DN 130:233066
TI Structure and polymorphism of the Chironomus thummi gene encoding special
lobe-specific silk protein, ssp160
AU Berezikov, Eugene; Blinov, Alexander G.; Scherbik, Svetlana; Cox, Carol
K.; Case, Steven T.
CS Laboratory of Cell Biology, Institute of Cytology and Genetics, Novosibirsk, 630090, Russia
SO Gene (1998), 223(1-2), 347-354
CODEN: GENED6; ISSN: 0378-1119
PB Elsevier Science B.V.

DT Journal
LA English
AB CDNA encoding Chironomus thummi ssp160 was used to isolate a genomic clone that hybridized in situ to band A2b on polytene chromosome IV, the site of the ssp160 gene. DNA sequencing, primer extension and gene/cDNA nucleotide **sequence alignment** revealed the gene contains six **exons** and five introns; 70% of ssp160 is encoded in exon 3. Variations between cDNA and gene sequences led to the design of a polymerase chain reaction, restriction fragment length polymorphism assay that was subsequently used to demonstrate the existence of polymorphic alleles whose distribution varied between geog. separated populations of larvae. The polymorphism is associated with codon deletions in a six-amino-acid repeat containing an N-linked glycosylation motif. These deletions may have resulted from slipped-strand mispairing during DNA replication.
RE.CNT 26 THERE ARE 26 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 14 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:783611 CAPLUS Full-text
DN 130:122813
TI Identification of a domain on the integrin α5 subunit implicated in
cell spreading and signaling
AU Cao, Zuojun; Huang, Kun; Horwitz, Alan F.
CS Department of Biochemistry, University of Illinois at Urbana-Champaign,
Urbana, IL, 61801, USA
SO Journal of Biological Chemistry (1998), 273(48), 31670-31679
CODEN: JBCHA3; ISSN: 0021-9258
PB American Society for Biochemistry and Molecular Biology
DT Journal
LA English
AB The α5β1 integrin is a cell surface receptor for fibronectin implicated in several cellular activities including cell proliferation, differentiation, and migration. The primary site at which the α5β1 integrin interacts with fibronectin is the RGD (Arg-Gly-Asp) amino acid sequence. In general, the sites on the integrin α subunits involved in ligand binding are not well characterized. Based on previous crosslinking studies, **sequence alignment**, predicted conformation, and intron-**exon** boundaries, the authors identified a 144-residue region (positions 223-367) on the α5 subunit as a putative binding region and divided it into four subdomains named domains I, II, III, and IV. Chimeric receptors were prepared in which sequences on the α5 subunit were exchanged with the corresponding sequences on the α6 subunit, which is specific for laminin and does not bind via an RGD sequence. The mutated human α5 integrin gene was transfected into CHO B2 cells, which are deficient in α5 expression. Only chimeras of domain III or IV express on the cell surface. Both of these chimeras decreased the adhesion, spreading, focal adhesion assembly, and migration on fibronectin. The adhesion of the chimeric receptors to fibronectin remained sensitive to the RGD peptide, and antibodies that inhibit interaction with the fibronectin synergy site and RGD loop remain inhibitory for the chimeras, indicating that our chimeras do not inhibit binding to either the RGD or synergy sites. Finally, the affinity of soluble fibronectin to cells via the α5β1 receptor decreased only about 3-fold. This decrease is substantially less than the observed effects on migration and spreading, which were not altered by changes in substrate concentration Thus, the alteration in binding sites does not easily account for the changes in cell spreading and focal adhesion assembly. The tyrosine phosphorylation and focal adhesion assembly that are seen when cells expressing the wild type α5 receptor adhere to fibronectin were inhibited in cells expressing the chimeric receptors. Therefore, our results suggest that the chimeras of these domains likely interrupt α5-mediated conformational signaling.
RE.CNT 61 THERE ARE 61 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 15 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:757502 CAPLUS Full-text
DN 130:105927
TI Structural organization of the human TOP2A and TOP2B genes
AU Lang, A. J.; Mirski, S. E. L.; Cummings, H. J.; Yu, Q.; Gerlach, J. H.;
Cole, S. P. C.
CS Department of Pharmacology and Toxicology, Queen's University, Kingston,
ON, K7L 3N6, Can.
SO Gene (1998), 221(2), 255-266
CODEN: GENED6; ISSN: 0378-1119
PB Elsevier Science B.V.
DT Journal
LA English
AB Eukaryotic topoisomerase II is an essential nuclear enzyme involved in processes such as chromosome condensation, chromatid separation, and in the relief of torsional stress that occurs during DNA transcription and replication. In cells from vertebrate species, there are two forms of the enzyme, designated α and β. Human topoisomerase IIα (TOP2A) is encoded by the TOP2A gene on chromosome 17q21-22, and human topoisomerase IIβ (TOP2B) is encoded by the TOP2B gene on chromosome 3p24. The protein products of these two genes are important cellular targets of several drugs widely used in the treatment of many human cancers, and a variety of mutations in TOP2A have been associated with the development of drug resistance. In the present study, we have defined the intron-exon structures of TOP2A and TOP2B. TOP2A is approx. 30 kb whereas TOP2B is at least 49 kb. TOP2A and TOP2B contain 35 and 36 exons, resp., and both genes contain a high proportion of class 0 introns. **Alignment** of the amino-acid **sequences** of the two proteins indicates that the intron-**exon** organization of the two genes is highly conserved, except for the regions encoding the extreme NH2 and COOH termini of the proteins. These findings suggest strongly that the vertebrate isoforms evolved by duplication of an ancestral gene. Mutations in TOP2A associated with drug resistance show clustering in exons 12, 13, 19-21 and 34-35. Knowledge of the genomic organization of TOP2A and TOP2B will be useful for detection of mutations in clin. samples from patients with drug-resistant malignant disease.
RE.CNT 38 THERE ARE 38 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 16 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:737306 CAPLUS Full-text
DN 130:91136
TI Divergent structures of Caenorhabditis elegans cytochrome P450 genes
suggest the frequent loss and gain of introns during the evolution of
nematodes
AU Gotoh, Osamu
CS Saitama Cancer Center Research Institute, Saitama, 362, Japan
SO Molecular Biology and Evolution (1998), 15(11), 1447-1459
CODEN: MBEVEO; ISSN: 0737-4038

PB Society for Molecular Biology and Evolution
DT Journal
LA English
AB The Caenorhabditis elegans genome contains more than 60 cytochrome P 450 (CYP) genes. The exon-intron organizations of all of the available and potentially active C. elegans CYP genes were inferred by a newly developed program for predicting protein-coding **exons** based on the **alignment** of a genomic DNA **sequence** and a protein profile. From the predicted amino acid sequences, all of the C. elegans CYP genes except one were classified into three groups, which were closely related to the mammalian drug-metabolizing P 450 gene families CYP2, CYP3, and CYP4. The gene structures were strikingly divergent within each group; 20, 10, and 5 unique gene organizations were identified among 40, 18, and 5 genes in the CYP2-, CYP3-, and CYP4-related groups, resp. The degrees of divergence in gene organization were strongly correlated with those in the amino acid sequences of encoding proteins, and the min. rate of change in an intron insertion site was estimated to be about 90 times less frequent than amino acid substitutions. Parsimonious analyses suggested that frequent loss and gain of introns has occurred during the evolution of CYP genes in each group after the divergence of nematodes, arthropods, and deuterostomia. Few, if any, incidents of intron sliding were evident, and a model that did not allow intron insertions was highly inconsistent with the observations. All of these findings are explained better by the intron-late view than by the intron-early view.
RE.CNT 43 THERE ARE 43 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 17 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:736924 CAPLUS Full-text
DN 130:91132
TI The exon structure of the human MAGP-2 gene. Similarity with the MAGP-1
gene is confined to two exons encoding a cysteine-rich region
AU Hatzinikolas, George; Gibson, Mark A.
CS Department of Pathology, University of Adelaide, Adelaide, 5005, Australia
SO Journal of Biological Chemistry (1998), 273(45), 29309-29314
CODEN: JBCHA3; ISSN: 0021-9258
PB American Society for Biochemistry and Molecular Biology
DT Journal
LA English
AB A cDNA for human microfibril-assocd. glycoprotein-2 (MAGP-2) was used to screen a human leukocyte genomic DNA library in EMBL-3 vector. One clone, clone H (10 kilobase pairs (kbp)), was isolated that contained most of the MAGP-2 gene. The remainder of the 3' end of the gene was obtained by direct polymerase chain reaction amplification of genomic DNA. The human MAGP-2 gene was found to be about 11 kbp in size and to contain 10 evenly distributed exons. The internal exons range in size from 30 base pairs (bp) to 88 bp with exons 4 and 6 the only exons of equal size (45 bp). All internal intron: exon junctions are defined by canonical splice donor and acceptor sites. Each junction has a 1/2 codon split with the exception of the exon 8/9 junction, which has a 2/1 split. The translation initiation codon is in exon 2, and the final exon contains 110 bp of coding sequence, including 2 cysteine codons. Primer extension expts. identified only one major transcription initiation site, 213 bases upstream of the ATG site. Rapid anal. of cDNA ends-polymerase chain reaction anal. of the 5' end of MAGP-2 mRNA from placenta confirmed this result and did not detect any alternative splicing of transcripts. The putative promoter region of the MAGP-2 gene was found to be AT-rich and it lacked a TATA box and other common regulatory elements. However the sequence surrounding the transcription start site CTCA(+1)TTCC was similar to the consensus CTCA(+1)NTCT (N is any nucleoside) for an initiator element found in terminal deoxynucleotidyltransferase and a number of other highly regulated genes. Comparison with the previously characterized human MAGP-1 gene showed that structural similarity was largely confined to the exact size, **sequence**, and junction **alignment** of the two penultimate **exons** which encode the first six of the seven cysteine residues that are precisely spaced in both proteins. The findings are consistent with the growing evidence that, although MAGP-1 and MAGP-2 are both intimately involved in the biol. of fibrillin-containing microfibrils, the MAGPs are structurally, functionally, and developmentally diverse proteins which share one characteristic cysteine-rich motif.
RE.CNT 37 THERE ARE 37 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 18 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:611708 CAPLUS Full-text
DN 130:11149
TI Identification of a new DPB1 allele (DPB1*7901) by sequence-based typing
AU Voorter, C.; Chatelain, B.; Sintnicolaas, K.; Tilanus, M.; Hidajat, M.;
Van Den Berg-Loonen, E.
CS Tissue Typing Laboratory, University Hospital Maastricht, Maastricht, 6202
AZ, Neth.
SO Tissue Antigens (1998), 52(2), 193-195
CODEN: TSANA2; ISSN: 0001-2815
PB Munksgaard International Publishers Ltd.
DT Journal
LA English
AB Sequence-based typing was used to detect the new HLA-DP antigen gene allele DPB1*7901. When the nucleotide **sequence** of the new allele was **aligned** with **exon** 2 **sequences** of several other DPB1 alleles, the highest homol. was with DPB1*2501 and *3701, both showing one nucleotide difference with the new allele. In both cases the nucleotide difference results in a leucine to isoleucine change at codon 65.
RE.CNT 7 THERE ARE 7 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 19 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:565016 CAPLUS Full-text
DN 129:271361
TI cDNA sequencing and analysis of POV1 (PB39): a novel gene up-regulated in
prostate cancer
AU Cole, Kristina A.; Chuaqui, Rodrigo F.; Katz, Kenneth; Pack, Svetlana;
Zhuang, Zhengping; Cole, Catherine E.; Lyne, John C.; Linehan, W. Marston;
Liotta, Lance A.; Emmert-Buck, Michael R.
CS Lab. Pathology, Division Clinical Sciences, National Cancer Institute,
National Institute Health, Bethesda, MD, 20892, USA
SO Genomics (1998), 51(2), 282-287
CODEN: GNMCEP; ISSN: 0888-7543
PB Academic Press
DT Journal
LA English
AB We recently identified a novel gene (PB39) (HGMW-approved symbol POV1) whose expression is up-regulated in human prostate cancer using tissue microdissection-

based differential display anal. In the present study we report the full-length sequencing of PB39 cDNA, genomic localization of the PB39 gene, and genomic sequence of the mouse homolog. The full-length human cDNA is 2317 nucleotides in length and contains an open reading frame of 559 amino acids which does not shown homol. with any reported human genes. The N-terminus contains charged amino acids and a helical loop pattern suggestive of an srp leader sequence for a secreted protein. Fluorescence in situ hybridization using PB39 cDNA as probe mapped the gene to chromosome 11p11.1-p11.2. Comparison of PB39 cDNA sequence with murine sequence available in the public database identified a region of previously sequenced mouse genomic DNA showing 67% amino acid sequence homol. with human PB39. Based on **alignment** and comparison to the human cDNA the mouse genomic **sequence** suggests that are at least 14 **exons** in the mouse gene spread over approx. 100 kb of genomic sequence. Further anal. of PB39 expression in human tissues shows the presence of a unique splice variant mRNA that appears to be primarily associated with fetal tissues and tumors. Interestingly, the unique splice variant appears in prostatic intraepithelial neoplasia, a microscopic precursor lesion of prostate cancer. The current data support the hypothesis that PB39 plays a role in the development of human prostate cancer and will be useful in the anal. of the gene product in further human and murine studies. (c) 1998 Academic Press.

RE.CNT 14 THERE ARE 14 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 20 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:310730 CAPLUS Full-text
DN 129:145546
TI Genomic organization and cDNA sequence of the rat RT1-DOb gene
AU Koda, Toshiaki; Kishi, Masahiko; Kakinuma, M.; Tank, W. R.
CS Institute of Immunological Science, Section of Bacterial Infection,
Hokkaido University, Kita-15, Sapporo, Kita-ku, 060, Japan
SO Immunogenetics (1998), 48(1), 67-72
CODEN: IMNGBK; ISSN: 0093-7711
PB Springer-Verlag
DT Journal
LA English
AB The genomic sequence of a DNA fragment contg. the LEC rat RT1-DOb gene was determined **Exon**/intron organization was defined by **aligning** the **sequence** with the mouse counterpart and a cDNA clone of Sprague-Dawley rat. The RT1-DOb gene consists of 6 exons and spans .apprx.7 kilobases. Sequences of the β1 domain-encoding region of the RT1-DOb gene from 22 rat strains revealed 6 alleles at the nucleotide level and 4 alleles at the amino acid level.

RE.CNT 18 THERE ARE 18 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 21 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:270783 CAPLUS Full-text
DN 129:53249
TI Gene structure, cDNA cloning, and expression of the rat cytokine-induced
neutrophil chemoattractant-2 (CINC-2) gene
AU Shibata, Futoshi; Konishi, Kiyoshi; Nakagawa, Hideo
CS Department of Physiological Chemistry, Faculty of Pharmaceutical Sciences,
Toyama Medical and Pharmaceutical University, Toyama, 930-01, Japan
SO Cytokine (1998), 10(3), 169-174
CODEN: CYTIE9; ISSN: 1043-4666
PB Academic Press Ltd.
DT Journal
LA English
AB Cytokine-induced neutrophil chemoattractant-2 (CINC-2) belongs to the CXC chemokine family and consists of two isoforms, CINC-2α and CINC-2β. The authors have studied the genomic organization and expression of the CINC-2 gene. The gene spans approx. 14 kb and is composed of three common exons, one CINC-2α-specific exon and two CINC-2β specific exons. This finding suggests that two isoforms of CINC-2 are encoded by mRNAs produced by alternative splicing. Each isoform is encoded in four **exons**, and **exon**-intron boundaries are placed identically within the **aligned sequences** of CXC chemokines. The CINC-2α-specific exon encodes an extra C-terminal serine residue, in addition to three amino acid residues (DKS) which were determined from amino acid sequence anal. of CINC-2α previously. The 5' flanking region of the gene contains a TATA box and putative binding sites for NF-κB and AP-1. Northern blot analyses showed that the mRNA level for CINC-2 was very low in rat peritoneal macrophages without stimulation and increased up to 4 h after lipopolysaccharide stimulation, similar to that for CINC-1 or CINC-3. Thereafter, the mRNA expression decreased gradually. However, the mRNA level of CINC-2 remained high 24 h after stimulation, in contrast to that of CINC-1 or CINC-3. These data indicate the expression of CINC-2 is regulated differently among the CINCs.

RE.CNT 28 THERE ARE 28 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 22 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:177187 CAPLUS Full-text
DN 128:279423
TI High sequence similarity within ras exons 1 and 2 in different mammalian
species and phylogenetic divergence of the ras gene family
AU Watzinger, F.; Mayr, B.; Haring, E.; Lion, T.
CS Children's Cancer Res. Inst., St. Anna Kinderspital, Vienna, A-1090,
Austria
SO Mammalian Genome (1998), 9(3), 214-219
CODEN: MAMGEC; ISSN: 0938-8990
PB Springer-Verlag New York Inc.
DT Journal
LA English
AB We have detd. the canine and feline N-, K-, and H-ras gene sequences from position +23 to +270 covering exons I and II which contain the mutational hot spot codons 12, 13, and 61. The results were used to assess the degree of similarity between ras gene DNA regions containing the critical domains affected in neoplastic disorders in different mammalian species. The comparative analyses performed included human, canine, feline, murine, rattine, and whenever possible, bovine, leporine (rabbit), porcelline (guinea pig), and mesocricetine (hamster) ras gene sequences within the region of interest. Comparison of feline and canine nucleotide sequences with the corresponding regions in human DNA revealed a sequence similarity greater than 85% to the human sequence. Contemporaneous anal. of previously published ras DNA sequences from other mammalian species showed a similar degree of homol. to human DNA. Most nucleotide differences observed represented synonymous changes without effect on the amino acid sequence of the resp. proteins. For assessment of the phylogenetic evolution of ras gene family, a maximum parsimony dendrogram based

on multiple **sequence alignment** of the common region of **exons** I and II in the N-, K-, and H-ras genes was constructed. Interestingly, a higher substitution rate among the H-ras genes became apparent, indicating accelerated sequence evolution within this particular clade. The most parsimonious tree clearly shows that the duplications giving rise to the three ras genes must have occurred before the mammalian radiation.

RE.CNT 30 THERE ARE 30 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 23 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:88264 CAPLUS Full-text
DN 128:226856
TI Comparative sequence analysis of a gene-rich cluster at human chromosome
12p13 and its syntenic region in mouse chromosome 6
AU Ansari-Lari, M. Ali; Oeltjen, John C.; Schwartz, Scott; Zhang, Zheng;
Muzny, Donna M.; Lu, Jing; Gorrell, James H.; Chinault, A. Craig; Belmont,
John W.; Miller, Webb; Gibbs, Richard A.
CS Department Molecular and Human Genetics, Baylor College Medicine, Houston,
TX, 77030, USA
SO Genome Research (1998), 8(1), 29-40
CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB The Human Genome Project has created a formidable challenge: the extn. of biol. information from extensive amts. of raw sequence. With the increasing availability of genomic sequence from other species, one approach to extracting coding and regulatory element information is through cross-species sequence comparison. To assess the strengths and weaknesses of this methodol. for large-scale sequence anal., 227 kb of mouse sequence syntenic to a gene-rich cluster on human chromosome 12p13 was obtained. Primarily through percent identity plots (PIPs) of SIM comparative **sequence alignments**, the **sequence** of coding reigns, putative alternative **exons**, conserved noncoding regions, and correlation in repetitive element insertions were easily determined The anal. demonstrated that the number, order, and orientation of all 17 genes are conserved between the two species, whereas two human pseudogenes are absent in mouse. In addition, apart from MIRs, no direct correlation of distribution or position of the majority of repetitive elements between the two species is seen. Finally, in examining the synonymous and nonsynonymous substitution rates in the conserved genes, a large variation in nonsynonymous rats is observed indicating that the genes in this region are diverging at different rates. This study indicates the utility and strength of large-scale cross-species sequence comparisons in the extraction of biol. information from raw sequence, especially when combined with other computational tools such as GRAIL and BLAST.

RE.CNT 43 THERE ARE 43 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 24 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:62613 CAPLUS Full-text
DN 128:163469
TI Characteristics of nucleotide sequences flanking the trans-spliced leader
SL1 exon in Dirofilaria immitis, Brugia malayi, and Brugia pahangi
AU Harasawa, Ryo; Maeda, Ryuichiro; Nogami, Sadao; Nakagaki, Kazuhide;
Yoshida, Motonobu; Kataoka, Yoko; Kobayashi, Hikari; Katae, Hiromi;
Hayashi, Yoshihiro
CS Animal Center Biomedical Research, Fac. Medicine, Univ. Tokyo, Tokyo, 113,
Japan
SO Journal of Veterinary Medical Science (1997), 59(12), 1149-1152
CODEN: JVMSEQ; ISSN: 0916-7250
PB Japanese Society of Veterinary Science
DT Journal
LA English
AB Nucleotide sequences surrounding the trans-spliced leader SL1 exon in the 5S rRNA gene spacer regions of Dirofilaria immitis, Brugia malayi, and B. pahangi were determined after PCR amplification, aligned with the genus Onchocerca for comparison, and used for the prediction of secondary structures. The nucleotide sequence of this region in B. pahangi was first shown in the present study. Hypothetical secondary structures of the spacer region suggested that the SL1 transcript is capable to form a stable stem-loop structure which may render transposition of the SL1 sequence to mRNA mols. A homologous sequence to Sm-binding site was assigned on a bulge loop. No significant difference was observed in adult worms of D. immitis irresp. of sex or location. No difference was apparent between the two species in genus Brugia.

RE.CNT 18 THERE ARE 18 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 25 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:46099 CAPLUS Full-text
DN 128:166194
TI Presence of solitary exon 1 sequences in the HLA-DR region
AU Gongora, Rafael
CS Abteilung Immungenetik, Max-Planck-Institut fur Biologie, Tubingen,
DE-72076, Germany
SO Hereditas (Lund, Sweden) (1997), 127(1-2), 47-49
CODEN: HEREAY; ISSN: 0018-0661
PB Mendelian Society of Lund
DT Journal
LA English
AB Conservation of the pseudogene DRB9 segment as opposed to high variability of the adjacent segment(s) was examined by **aligning** all known **exon** 1 **sequences** of human DRB genes and solitary **exon** 1 **sequences** (S1, S3, S4, and S5) upstream of DRB9. The S3 solitary exon 1 downstream of the DRB5 locus (haplotype DR15) was sequenced. From anal. of flanking introns, the pseudogene S3, the solitary exon 1 nearest DRB9, does not represent exon 1 of DRB9, which now consists of a solitary exon 2. The origins of pseudogenes S1, S3, S4, and S5, representing exons 1 on areas upstream of DRB9 are considered.

RE.CNT 12 THERE ARE 12 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 26 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1997:784782 CAPLUS Full-text
DN 128:98532
TI A tool for analyzing and annotating genomic sequences
AU Huang, Xiaoqiu; Adams, Mark D.; Zhou, Hao; Kerlavage, Anthony R.
CS Department Computer Science, Michigan Technological University, Houghton,
MI, 49931, USA

SO Genomics (1997), 46(1), 37-45
CODEN: GNMCEP; ISSN: 0888-7543
PB Academic Press
DT Journal
LA English
AB We describe a tool for analyzing and annotating large genomic sequences containing introns. The anal. and annotation tool (AAT) includes two sets of programs, one for comparing the query sequence with a protein database and the other for comparing the query with a cDNA database. Each set contains a fast database search program and a rigorous alignment program. The database search program quickly identifies regions of the query sequence that are similar to a database sequence. Then the alignment program constructs an optimal alignment for each region and the database sequence. The **alignment** program also reports the coordinates of **exons** in the query **sequence.** Pairwise alignments of the query sequence with protein and cDNA database sequences are combined into multiple sequence alignments, which provide a view of all protein and cDNA sequence matching a query region. On a data set of 570 DNA sequences, AAT identified 94% of coding nucleotides correctly and 74% of exons exactly. Results of analyzing a human BAC sequence with the AAT tool are also presented. The AAT tool reduces the labor-intensive work of locating the exons of the query sequence and improves the process of defining intron-exon boundaries by using the wealth of available protein and cDNA data.
RE.CNT 19 THERE ARE 19 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 27 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1997:641741 CAPLUS Full-text
DN 127:315360
TI Intra- and inter-individual heterogeneity in exon 2 of the MDR1 gene in
primary breast carcinoma and healthy individuals
AU Aalto, Yan; Teglund, Stephen; Andersson, Ulrika; Blanco, Guillermo;
Hammarstrom, Sten; Henriksson, Roger
CS Department Oncology, Umea University, Umea, S-901 85, Swed.
SO International Journal of Oncology (1997), 11(4), 697-701
CODEN: IJONES; ISSN: 1019-6439
PB International Journal of Oncology
DT Journal
LA English
AB Increased expression of P-glycoprotein, encoded by the MDR1 gene, is considered to be responsible for chemotherapy failure in a number of human cancers. Although it is clear that mutations in the MDR1 gene affect substrate specificity of the transporter in multidrug-resistant cell lines, scant interest has been directed at whether mutations have a unique clin. presentation. To address this question, exon 2 of the MDR1 gene was studied in 9 patients with primary breast carcinoma and 9 healthy controls using PCR and DNA sequence anal. To reduce the possibility of nucleotide misincorporations introduced by Taq polymerase, sequencing of six subclones of each DNA specimen was performed. A mutation was seen as a substitution from G to A at position -1 in two patients and one control. An A to G nucleotide substitution giving rise to an amino acid substitution (Asn→Asp) in codon 21 at the first potential N-glycosylation site of the P-glycoprotein was seen in primary tumors from four patients and in an axillar lymph node metastases from one of these patients. This mutation was also seen in two healthy individuals, which similar to the patients, both seem to be heterozygous for this MDR1 exon 2 allele. Three other mutations were also found in the patients; a substitution of A to G at position 23 and A to G at position 52 in the same patient and in another patient, G at position 42 was changed to A. However, the last three mutations were not confirmed by repeating anal. of the original genomic sample. The results revealed different distribution of a point mutation between various parts of the same primary tumor and between a lymph node metastasis and the primary tumor tissue. Thus, demonstrating both intra- and inter-tumor heterogeneity. The results also emphasized constitutional allelic variation in the MDR1 gene. Whether this might affect sensitivity to chemotherapy has to be further evaluated.
RE.CNT 33 THERE ARE 33 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 28 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1997:603027 CAPLUS Full-text
DN 127:273727
TI Characterization and targeting of the murine α2-antiplasmin gene
AU Okada, K.; Lijnen, H. R.; Dewerchin, M.; Belayew, A.; Matsuo, O.; Collen,
D.; Bernaerts, R.
CS Center Molecular Vascular Biology, University Leuven, Louvain, B-3000,
Belg.
SO Thrombosis and Haemostasis (1997), 78(3), 1104-1110
CODEN: THHADQ; ISSN: 0340-6245
PB Schattauer
DT Journal
LA English
AB α2-Antiplasmin (α2-AP) is the main physiol. plasmin inhibitor in mammalian plasma. As a 1st step toward the generation of α2-AP deficient mice, the murine α2-AP gene was characterized and a targeting vector for homologous recombination in embryonic stem (ES) cells constructed. **Alignment** of nucleotide **sequences** obtained from genomic subclones allowed location of **exons** 2 through 10 of the α2-AP gene, but failed to identify the 5' boundary of exon 1. Compared to the human gene, exons 2 through 9 in the murine gene have identical size and intron-exon boundaries obeying the GT/AG rule. The 5' boundary of exon 10 is identical in both genes while the 3' non-coding region is 64 bp longer in the human gene. Introns 2, 3, 6, and 8 have similar sizes in the mouse and human genes; intron 1 is 6-fold smaller, introns 5, 7, and 9 are 2-3-fold smaller, whereas intron 4 is about 2-fold larger in the mouse gene. Compared to the human 5' flanking sequence, an insertion of a simple repeat region with sequence (TGG)n has occurred. The open reading frame of the mouse α2-AP gene encodes a 491-amino-acid protein comprising the exptl. determined NH2-terminus of the mature protein Val-Asp-Leu-Pro-Gly-. A targeting vector, pPNT.α2-AP, was constructed by introducing a homologous sequence of 8.3 kb in total in the parental pPNT vector. In pPNT.α2-AP, the neomycin resistance expression cassette replaces a 7 kb genomic fragment comprising exon 2 through part of exon ID (including the stop codon), which represents the entire sequence encoding the mature protein, including the fibrin-binding domain, the reactive site peptide bond and the plasmin(ogen)-binding region. Electroporation of 129R1 embryonic stem (ES) cells with the linearized vector pPNT.α2-AP yielded 3 targeted clones with correct homologous recombination at the 5'- and 3'-ends, as confirmed by Southern blot anal. of purified genomic DNA with appropriate restriction enzymes and probes. These

targeted clones will be used to generate α2-AP deficient mice.

L6 ANSWER 29 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1997:131758 CAPLUS Full-text
DN 126:196556
TI LALNVIEW: a graphical viewer for pairwise sequence alignments
AU Duret, Laurent; Gasteiger, Elisabeth; Perriere, Guy
CS Department of Medical Biochemistry, University of Geneva, Geneva,
CH-1211/4, Switz.
SO Computer Applications in the Biosciences (1996), 12(6), 507-510
CODEN: COABER; ISSN: 0266-7061
PB Oxford University Press
DT Journal
LA English
AB LALNVIEW is a graphical program for visualizing local alignments between two sequences (protein or nucleic acids). Sequences are represented by colored rectangles to give an overall picture of their similarities. LALNVIEW can display **sequence** features (**exon**, intron, active site, domain, propeptide, etc.) along with the **alignment.** When using LALNVIEW through our Web servers, sequence features are automatically extracted from database annotations (SWISS-PROT, GenBank, EMBL or HOVERGEN) and displayed with the alignment. LALNVIEW is a useful tool for analyzing pairwise sequence alignments and for making the link between sequence homol. and what is known about the structure or function of sequences. LALNVIEW executables for UNIX, Macintosh and PC computers are freely available from our server (http://expasy.hcuge.ch/sprot/lalnview.ht ml).

L6 ANSWER 30 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:706561 CAPLUS Full-text
DN 126:15226
TI Naturally occurring nucleosome positioning signals in human exons and
introns
AU Baldi, Pierre; Brunak, Soeren; Chauvin, Yves; Krogh, Anders
CS Div. Biol., California Inst. Technol., Pasadena, CA, 91125, USA
SO Journal of Molecular Biology (1996), 263(4), 503-510
CODEN: JMOBAK; ISSN: 0022-2836
PB Academic
DT Journal
LA English
AB We describe the structural implications of a periodic pattern found in human exons and introns by hidden Markov models. We show that exons (besides the reading frame) have a specific sequential structure in the form of a pattern with triplet consensus non-T(A/T)G, and a minimal periodicity of roughly ten nucleotides. The periodic pattern is also present in intron sequences, although the strength per nucleotide is weaker. Using two independent profile methods based on triplet bendability parameters from DNase I expts. and nucleosome positioning data, we show that the pattern in multiple **alignments** of internal **exon** and intron **sequences** corresponds to a periodic "in phase" bending potential towards the major groove of the DNA. The nucleosome positioning data show that the consensus triplets (and their complements) have a preference for locations on a bent double helix where the major groove faces inward and is compressed. The in-phase triplets are located adjacent to TCC/GGC triplets known to have the strongest bias in their positioning on the nucleosome. Anal. of mRNA sequences encoding proteins with known tertiary structure exclude the possibility that the pattern is a consequence of the previously well-known periodicity caused by the encoding of alpha-helixes in proteins. Finally, we discuss the relation between the bending potential of coding and non-coding regions and its impact on the translational positioning of nucleosomes and the recognition of genes by the transcriptional machinery.

L6 ANSWER 31 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:681925 CAPLUS Full-text
DN 126:2060
TI Genomic organization and DNA sequences of two human phenol sulfotransferase genes (STP1 and STP2) on the short arm of chromosome 16
AU Dooley, Thomas P.; Huang, Zimei
CS Molecular Pharmacology, Southern Res. Inst., Birmingham, AL, 35205, USA
SO Biochemical and Biophysical Research Communications (1996), 228(1),
134-140
CODEN: BBRCA9; ISSN: 0006-291X
PB Academic
DT Journal
LA English
AB A family of human phenol sulfotransferase genes has been suggested by the cloning of numerous cDNA isolates from different tissues. The STM gene encoding the monoamine neurotransmitter-preferring sulfotransferase, M-PST, and a portion of the STP1 gene encoding the phenol-preferring isoenzyme, P-PST, were previously cloned and sequenced. Both genes were mapped to a small region on the short arm of chromosome 16. This report describes the sequencing and genomic organization of the STP1 and STP2 genes from a single cosmid clone obtained from chromosome 16p12.1-p11.2. STP1 and STP2 are 95.9% identical at the amino acid sequence level, whereas the STM gene is only 92.9% and 90.5% identical to STP1 and STP2, resp. **Alignment** of the genomic **sequences** indicated that all three genes have 7 coding **exons** and conserved intron-exon boundaries. These results facilitated the assignment of previously published cDNA isolates as "alleles" of the individual STM, STP1, and STP2 loci on 16p, and provide a greater understanding of the complexity and roles of the phenol sulfotransferase gene family in the metabolism of endogenous and xenobiotic agents.

L6 ANSWER 32 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:329954 CAPLUS Full-text
DN 125:27061
TI Detection and identification of human pathogenic Leishmania and
Trypanosoma species by hybridization of PCR-amplified mini-exon repeats
AU Ramos, Anthea; Maslov, Dmitri A.; Fernandes, Octavio; Campbell, David A.;
Simpson, Larry
CS Howard Hughes Medical Institute, University California, Los Angeles, CA,
90095-1662, USA
SO Experimental Parasitology (1996), 82(3), 242-250
CODEN: EXPAAA; ISSN: 0014-4894
PB Academic
DT Journal
LA English
AB Detection and identification of human pathogenic Leishmania and Trypanosoma species by hybridization of PCR-amplified mini-exon repeats. A single pair of PCR primers within a conserved region of the mini-exon repeat was used to amplify the repeats from 10 species of

pathogenic Leishmania belonging to four major clin. groups and also from three species of Trypanosoma. Oligonucleotide hybridization probes for the detection and identification of the PCR-amplified repeats were constructed from **alignments** of mini-**exon** intron and intergenic **sequences.** The probes generated from mini-exon intergenic regions of the L. (V.) braziliensis, L. (L.) donovani, and L. (L.) mexicana species hybridized specifically to their cognate groups without discriminating between the species within the groups. The probes for L. (L.) major and L. (L.) aethiopica were species-specific, while the L. (L.) tropica probe also hybridized with the L. (L.) aethiopica mini-exon repeat. The mini-exon intron-derived probes for T. cruzi, T. rangeli, and T. brucei were species-specific. This method involving the detection of specific PCR-amplified products produced using a single primer set represents a novel sensitive and specific assay for multiple trypanosomatid species and groups.

L6 ANSWER 33 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:82006 CAPLUS Full-text
DN 124:137523
TI The role of PRP8 protein in nuclear pre-mRNA splicing in yeast
AU Beggs, Jean D.; Teigelkamp, Stefan; Newman, Andrew J.
CS Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh, EH9 3JR, UK
SO Journal of Cell Science, Supplement (1995), 19, 101-5
CODEN: JCSSEP; ISSN: 0269-3518
DT Journal
LA English
AB The removal of introns from precursor mRNAs occurs in a large complex, the spliceosome, that contains many proteins and five small nuclear RNAs (snRNAs). The snRNAs interact with the intron-containing substrate RNA and with each other to form a dynamic network of RNA interactions that define the intron and promote splicing. There is evidence that protein splicing factors play important roles in regulating RNA interactions in the spliceosome. PRP8 is a highly conserved protein that is associated in particles with the U5 snRNA and directly binds the substrate RNA in spliceosomes. UV crosslinking has been used to map the binding sites, and shows extensive interaction between PRP8 protein and the 5' exon prior to the first step of splicing and with the 3' splice site region subsequently. It is proposed that PRP8 protein may stabilize fragile interactions between the U5 snRNA and **exon sequences** at the splice sites, to anchor and **align** them in the catalytic center of the spliceosome.

L6 ANSWER 34 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:75722 CAPLUS Full-text
DN 124:195371
TI Genomic organization of the human tissue inhibitor of metalloproteinases-3 (TIMP3)
AU Stoehr, Heidi; Roomp, Kirsten; Felbor, Ute; Weber, Bernhard H. F.
CS Inst. Humangenetik, Univ. Wuerzburg, Wuerzburg, D-97074, Germany
SO Genome Research (1995), 5(5), 483-7
CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB Recently, the authors have shown that mutations in TIMP3 cause the autosomal dominant disorder Sorsby's fundus dystrophy. This is a macular degeneration disorder with characteristic extracellular matrix irregularities in Bruch's membrane. To further facilitate mutational anal. and to provide a basis for functional studies, the authors report the genomic organization of the human TIMP3 gene. **Alignment** of the genomic **sequences** to the published TIMP3 cDNAs revealed the **exon**/intron organization of the human TIMP3 gene that is encoded by 5 exons with the most likely assignment of donor and acceptor splice junctions following the 5'-GT-AG-3' rule. Exon 1 contains the translation initiation start codon ATG as well as 280 bp of upstream sequence corresponding to the most 5'-extending cDNA isolated. This suggests that the 5'-flanking region of the TIMP3 gene is not interrupted further by intervening sequences. The overall organization of the human TIMP3 gene seems very similar to that of the recently reported murine homolog.

L6 ANSWER 35 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:29715 CAPLUS Full-text
DN 124:108379
TI Sequence and evolution of the Drosophila pseudoobscura glycerol-3-phosphate dehydrogenase locus
AU Wells, R. Spencer
CS Museum Comparative Zool. Lab., Harvard Univ., Cambridge, MA, 02138, USA
SO Journal of Molecular Evolution (1995), 41(6), 886-93
CODEN: JMEVAU; ISSN: 0022-2844
PB Springer
DT Journal
LA English
AB The Gpdh genomic region has been cloned and sequenced in Drosophila pseudoobscura. A total of 6.8 kb of sequence was obtained, encompassing all eight exons of the gene. The **exons** have been **aligned** with the **sequence** from D. melanogaster, and the rates of synonymous and nonsynonymous substitution have been compared to those of other genes sequenced in these two species. Gpdh has the lowest rate of nonsynonymous substitution yet seen in genes sequenced in both D. pseudoobscura and D. melanogaster. No insertion/deletion events were observed, and the overall architecture of the gene (i.e., intron sites, etc.) is conserved. An interesting amino acid reversal was noted between the D. melanogaster Fast allele and the D. pseudoobscura gene.

L6 ANSWER 36 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:14837 CAPLUS Full-text
DN 124:108290
TI Tissue-dependent expression of a novel splice variant of the human estrogen receptor
AU Daffada, Angela I.; Dowsett, Mitchell
CS Inst. of Cancer Research, Royal Marsden Hospital, London, SW3 6FF, UK
SO Journal of Steroid Biochemistry and Molecular Biology (1995), 55(3/4), 413-21
CODEN: JSBBEZ; ISSN: 0960-0760
PB Elsevier
DT Journal
LA English
AB The authors have isolated a novel splice variant of ER mRNA from normal endometrial tissue using RT/PCR. The variant contains an unusual splice junction formed by splicing sequences within exons 4 and 7 together. The translated protein product would be predicted to lack part of exon 4, all of exons 5 and 6 and, due to a missense **alignment** at the new splice junction, the remaining **sequence** from **exon** 7 would be translated out of frame

and terminate at the exon 7/8 splice junction. As a result, the protein would lack most of the hormone binding domain (HBD) and the major estrogen-dependent transactivating region (AF-2), but still contain the DNA binding domain (DNA-BD) and N-terminal transactivating region (AF-1). In contrast to the exon 5 deleted variant of ER (Δ5), which was expressed in both normal endometrium and liver, this novel variant was present in endometrium but not in liver samples. These results confirm that some ER splice variants are expressed in normal, non-malignant estrogen responsive tissues. In addition, they demonstrate the tissue specific expression of a novel and interesting splice variant of ER in these normal tissues.

L6 ANSWER 37 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1995:825723 CAPLUS Full-text
DN 123:331664
TI Molecular evolution of the genes encoding receptor tyrosine kinase with
immunoglobulin-like domains
AU Rousset, Dominique; Agnes, Francois; Lachaume, Philippe; Andre, Catherine;
Galibert, Francis
CS Laboratorie Biochimie Mol., Faculte Medecine, Rennes, 35043, Fr.
SO Journal of Molecular Evolution (1995), 41(4), 421-9
CODEN: JMEVAU; ISSN: 0022-2844
PB Springer
DT Journal
LA English
AB Receptor tyrosine kinases (RTK) with five, three, or seven Ig-like domains in their extracellular regions are classified as subclasses III, IV, and V, resp. Conservation of the exon/intron structure of the downstream part of the human KIT, FMS, and FLT3 genes that encode RTK of subclass III together with the particular chromosomal localization of these genes suggests that RTKIII genes have evolved from a common ancestor by cis and trans duplications. To strengthen this model of evolution and to determine if it can be extended to RTKIV and V genes, the authors constructed a phylogenetic tree of RTKII, IV, and V on the basis of a multiple **alignment** of their catalytic tyrosine kinase domain **sequence** and determined the **exon**/intron structure of PDGFRA (subclass III), FGFR4 (subclass IV), and FLT4 (subclass V) genes in their downstream parts. Phylogenetic analyses with amino acid or nucleotide sequences both resulted in one most parsimonious tree. The phylogenetic tress obtained indicate that all three subclasses are well individuated and that RTKII and RTKV are closer to each other than RTKIV. Furthermore, RTKIII and FLT4 (subclass V) genes possess the same exon/intron structure in their downstream part while the structure of the RTKIV genes is very similar to that of RTKIII and FLT4. Both approaches are in complete agreement and indicate that RTKIII, IV, and V genes most probably evolved from a common ancestor already "in pieces: by successive duplications involving entire genes.

L6 ANSWER 38 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1995:794154 CAPLUS Full-text
DN 123:337023
TI The identification and cloning of a murine major basic protein gene
expressed in eosinophils
AU Larson, Kirsten A.; Horton, Margaret A.; Madden, Benjamin J.; Gleich,
Gerald J.; Lee, Nancy A.; Lee, James J.
CS Dep. Biochemistry and Molecular Biology, Mayo Clinic, Scottsdale, AZ,
85259, USA
SO Journal of Immunology (1995), 155(6), 3002-12
CODEN: JOIMA3; ISSN: 0022-1767
PB American Association of Immunologists
DT Journal
LA English
AB The existence of a murine homolog of the major basic protein (MBP) found in human eosinophil granules was initially hypothesized from structural similarities at the electron microscopic level. The results presented in this study have extended these observations by describing the identification/purification of a mouse MBP (mMBP) and the cloning of the gene encoding this eosinophil granule protein. Using protein purification methodologies with extravascular eosinophils, an mMBP homolog has been identified on the basis of strong (64%) N-terminal sequence homol. with the mature human MBP (hMBP). Since hMBP results from a proteolytic cleavage of a precursor mol., this sequence conservation suggests that the mouse granule protein is processed by a similar mechanism. The gene encoding mMBP was isolated using a hMBP cDNA clone as a heterologous probe in low criteria screens of mouse genomic and cDNA libraries. The genomic structure and nucleotide **sequence** of the mMBP **exons** are well conserved with the human gene, although homol. **alignments** of the encoded proteins show that extensive sequence conservation occurs only in the mature portion of the MBP mols. Expression data demonstrate that this gene is transcriptionally active in tissues containing eosinophil progenitor cells, such as femoral bone marrow. Genomic Southern blots using the mMBP gene at reduced stringency reveal the potential existence of a second, more divergent MBP-like sequence in the mouse. This suggests that, as with guinea pigs, the mouse genome may also encode the eosinophil major basic protein from more than one gene.

L6 ANSWER 39 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1995:788541 CAPLUS Full-text
DN 123:332912
TI The 3'-terminal exon of the family of steroid and phenol sulfotransferase
genes is spliced at the N-terminal glycine of the universally conserved
GXXGXXK motif that forms the sulfonate donor binding site
AU Chiba, H.; Komatsu, K.; Lee, Y. C.; Tomizuka, T.; Strott, C. A.
CS Endocrinology and Reproduction Research Branch, National Inst. Health,
Bethesda, MD, 20892-4510, USA
SO Proceedings of the National Academy of Sciences of the United States of
America (1995), 92(18), 8176-9
CODEN: PNASA6; ISSN: 0027-8424
PB National Academy of Sciences
DT Journal
LA English
AB The guinea pig estrogen sulfotransferase gene has been cloned and compared to three other cloned steroid and phenol sulfotransferase genes (human estrogen sulfotransferase, human phenol sulfotransferase, and guinea pig 3α-hydroxysteroid sulfotransferase). The four sulfotransferase genes demonstrate a common outstanding feature: the splice sites for their 3'-terminal exons are identically located. I.e., the 3'-terminal exon splice sites involve a glycine that constitutes the N-terminal glycine of an invariably conserved GXXGXXK motif present in all steroid and phenol sulfotransferases for which primary structures are known. This consistency strongly suggests

that all steroid and phenol sulfotransferase genes will be similarly spliced. The GXXGXXK motif forms the active binding site for the universal sulfonate donor 3'-phosphoadenosine 5'-phosphosulfate. Amino acid sequence alignment of 19 cloned steroid and phenol sulfotransferases starting with the GXXGXXK motif indicates that the 3'-terminal exon for each steroid and phenol sulfotransferase gene encodes a similarly sized C-terminal fragment of the protein. Interestingly, on further anal. of the alignment, three distinct amino acid sequence patterns emerge. The presence of the conserved functional GXXGXXK motif suggests that the protein domains encoded by steroid and phenol sulfotransferase 3'-terminal exons have evolved from a common ancestor. Furthermore, it is hypothesized that during the course of evolution, the 3'-terminal exon further diverged into at least three sulfotransferase subdivisions: a phenol or aryl group, an estrogen or phenolic steroid group, and a neutral steroid group.

L6 ANSWER 40 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1993:576135 CAPLUS Full-text
DN 119:176135
TI A bacterial protein has homology with human chorionic gonadotropin (hCG)
AU Grover, Sanjeev; Woodward, Scott R.; Odell, William D.
CS Sch. Med., Univ. Utah, Salt Lake City, UT, 84132, USA
SO Biochemical and Biophysical Research Communications (1993), 193(3), 841-7
CODEN: BBRCA9; ISSN: 0006-291X
DT Journal
LA English
AB Studies have demonstrated the presence of a 48.5 kD cell wall protein in the bacterium, Xanthomonas maltophilia, which immunol. resembles the beta subunit of human chorionic gonadotropin. Primers were designed from the amino acid sequences of enzymically cleaved peptide fragments of this protein. These primers were used to obtain PCR amplified products, which were subsequently cloned in a PCR11TA cloning vector, and a 492 base pair nucleotide sequence was obtained with a 164 amino acid open reading frame. When this nucleotide **sequence** was **aligned** with **exon** 2 of genes 5 and 6 of the βhCG gene, a 53% homol. was observed The translated protein sequence had a 35% homol. with hCG and a 25% homol. with human LH.

L6 ANSWER 41 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1993:532683 CAPLUS Full-text
DN 119:132683
TI The human gene encoding tryptophanyl-tRNA synthetase: interferon-response
elements and exon-intron organization
AU Frolova, Lyudmila Y.; Grigorieva, Arina Y.; Sudomoina, Marina A.;
Kisselev, Lev L.
CS Engelhardt Inst. Mol. Biol., Moscow, 117984, Russia
SO Gene (1993), 128(2), 237-45
CODEN: GENED6; ISSN: 0378-1119
DT Journal
LA English
AB The cDNA encoding human tryptophanyl-tRNA synthetase (hWRS) has recently been cloned and sequenced. Independently, it has been shown that this protein is induced by interferons (IFN) γ and α. This unusual feature of a housekeeping enzyme raises the problem of how the gene is regulated. Since at present the genomic structure of hWRS is unknown, this issue remains unsolved. The exon-intron organization of hWRS was now deciphered. This gene consists of ≥12 exons that span >35 kb of DNA. At least 2 alternative noncoding exons precede 10 coding exons. Upstream from the first exon, two GGAAAN(N/-)GAAA sequences, which are considered to be IFN-stimulating response elements (ISRE), were revealed. The same consensus was also found in the intron region in close vicinity to the 5' end of the second exon. Thus, the IFN-stimulated synthesis of hWRS is presumably due to gene activation at the transcriptional level. **Alignment** of hWRS amino acid **sequences** showed that **exons** V-XI of hWRS encode regions of structural similarity with bacterial WRS, whereas the N-terminal portion of the protein encoded by exons II-IV exhibits no homol. with bacterial WRS. The enzymically active core enzyme generated by limited proteolysis is presumably encoded by exons V-XI. It is concluded that mammalian WRS is composed of 2 structurally and functionally different domains encoded by the 5' and 3' portions of its gene.

L6 ANSWER 42 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1993:487225 CAPLUS Full-text
DN 119:87225
TI Alternatively spliced mRNAs for human endothelin-2 and their tissue
distribution
AU O'Reilly, G.; Charnock-Jones, D. S.; Morrison, J. J.; Cameron, I. T.;
Davenport, A. P.; Smith, S. K.
CS Clin. Sch., Univ. Cambridge, Cambridge, UK
SO Biochemical and Biophysical Research Communications (1993), 193(3), 834-40
CODEN: BBRCA9; ISSN: 0006-291X
DT Journal
LA English
AB The cDNA for endothelin-2 (ET-2) has been previously cloned and characterized; however, ET-2 remains the least studied of the endothelin isopeptides and little is known of its function and location. In the present study, reverse transcriptase-polymerase chain reaction revealed the presence of 7 alternatively spliced mRNA variants encoding ET-2, with a specific pattern of distribution in various human tissues. Computer **alignment** and anal. of the DNA **sequences** demonstrated alternative splicing of 5 **exons** of 52, 169, 123, 99, and 174 base pairs, in the carboxy terminal region of the mRNA encoding preproET-2. This region contains sites for the post-transcriptional processing of preproET-2 into mature ET-2, therefore post-transcriptional processing may be disrupted or altered in these variants.

L6 ANSWER 43 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1993:229005 CAPLUS Full-text
DN 118:229005
TI Structural organization of the gene for prostaglandin D synthase in the
rat brain
AU Igarashi, Makoto; Nagata, Akihisa; Toy, Hiroyuki; Urade, Yoshihiro;
Hayaishi, Osamu
CS Res. Inst. Microbial Dis., Osaka Univ., Suita, 565, Japan
SO Proceedings of the National Academy of Sciences of the United States of
America (1992), 89(12), 5376-80
CODEN: PNASA6; ISSN: 0027-8424
DT Journal
LA English

AB A 3-kilobase-pair gene for rat brain prostaglandin D synthase [(5Z,13E)-(15S)-9α,11α-epidoxy-15-hydroxyprosta-5,13-dienoate D-isomerase, EC 5.3.99.2], which belongs to the lipocalin family, was isolated from a rat genomic DNA library by plaque hybridization with the cDNA for the enzyme. The gene contains seven exons, and all the splice donor and acceptor sites conform to the GT/AG rule. Transcription initiates at a guanine residue 39 base pairs upstream of the translation initiation codon, as determined by primer-extension anal. of rat brain mRNA. The 5'-flanking region of the gene lacks typical transcriptional regulatory sequences, such as TATA and CAAT boxes, but contains several sets of inverted repeats, direct repeats, and sequences resembling the transcriptional factor Sp1-binding site. The gene structure of prostaglandin D synthase is remarkably analogous to those of other lipocalins, such as β-lactoglobulin, α2-urinary globulin, placental protein 14, and α1-microglobulin, in terms of number and sizes of exons and phase of splicing of introns. Furthermore, in a multiple **alignment** of the deduced amino acid **sequences**, positions of **exon**/intron junction of the prostaglandin D synthase gene are highly conserved and located around the positions of those of the genes for other lipocalins despite a weak homol.

L6 ANSWER 44 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1993:226849 CAPLUS Full-text
DN 118:226849
TI Mouse apolipoprotein AI: cDNA-derived primary structure, gene organization and complete nucleotide sequence
AU Stoffel, Wilhelm; Mueller, Rolf; Binczek, Erika; Hofmann, Kay
CS Inst. Biochem., Med. Fak., Univ. Koeln, Cologne, Germany
SO Biological Chemistry Hoppe-Seyler (1992), 373(4), 187-93
CODEN: BCHSEI; ISSN: 0177-3593
DT Journal
LA English
AB A full-length apo AI-specific mouse liver cDNA clone was isolated with the human cDNA (892 bp) and the derived amino acid sequence coding a polypeptide of 264 residues described. The sequence showed a 70.7% homol. to the rat and 66% to the human apo AI sequence. With this cDNA as probe, the mouse apo AI gene was isolated and its organization analyzed. Four **exons**, 3 of which are coding **sequences**, are **aligned** similarly to the human gene. The gene embraces 1825 bp between the transcription start, and the poly(A)+ tail attached 62 bp downstream of the stop codon. The complete nucleotide sequence of the 4 exons and 3 introns of the mouse apo AI gene was determined and its homol. compared with that of the rat and human gene. Extensive deletions and a strongly reduced homol. of the 3 introns of the 2 genes are obvious.

L6 ANSWER 45 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1992:466026 CAPLUS Full-text
DN 117:66026
TI Multiple alignment using simulated annealing: branch point definition in human mRNA splicing
AU Lukashin, Alexander V.; Engelbrecht, Jacob; Brunak, Soeren
CS Dep. Phys. Chem., Tech. Univ. Denmark, Lyngby, DK-2800, Den.
SO Nucleic Acids Research (1992), 20(10), 2511-16
CODEN: NARHAD; ISSN: 0305-1048
DT Journal
LA English
AB A method for the simultaneous alignment of a very large no. of sequences using simulated annealing is presented. The total running time of the algorithm does not depend explicitly on the number of sequences treated. The method has been used for the simultaneous **alignment** of 1462 human intron **sequences** upstream of the intron-**exon** boundary. The consensus sequence of the aligned set together with a calcn. of the Shannon information clearly shows that several sequence motives are conserved: (1) a previously undetected guanosine rich region, (2) the branch point and (3) the polypyrimidine tract. The nucleotide frequencies at each position of the branch point consensus sequence qual. reproduce the frequencies of the exptl. determined branch points.

L6 ANSWER 46 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1992:446364 CAPLUS Full-text
DN 117:46364
TI Sequence and structural relationships in the cytokine family
AU Manavalan, Parthasarathy; Swope, Deborah L.; Withy, Raymond M.
CS Genzyme Corp., Framingham, MA, 01901, USA
SO Journal of Protein Chemistry (1992), 11(3), 321-31
CODEN: JPCHD2; ISSN: 0277-8033
DT Journal
LA English
AB The sequences of 9 different cytokines, growth hormone, and prolactin were aligned and their secondary structure predicted. The **alignment** reveals that each **exon** has a characteristic **sequence** pattern shared by all cytokines. The most striking sequence similarity is observed in exon 4, where the residue pair Phe-Leu is conserved in many cytokines. In addition, there are discrete homologous regions between two specific growth factors, including a high degree of homol. between granulocyte-macrophage colony-stimulating factor (GM-CSF) and interleukin 3 (IL-3). The secondary structure anal. predicts that exon 3 of all cytokines has an antiparallel helix-turn-helix motif, which is likely to form the central helical segments of a 4 α-helical bundle-type structure. Based on the secondary structure and the disulfide-bonding pattern, the topol. connectivity for a number of cytokines was predicted.

L6 ANSWER 47 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1992:441638 CAPLUS Full-text
DN 117:41638
TI The barley genes Ac11 and Ac13 encoding acyl carrier proteins I and III are located on different chromosomes
AU Hansen, Lars; Von Wettstein-Knowles, Penny
CS Dep. Physiol., Carlsberg Lab., Copenhagen Valby, DK-2500, Den.
SO Molecular and General Genetics (1991), 229(3), 467-78
CODEN: MGGEAE; ISSN: 0026-8925
DT Journal
LA English
AB Acyl carrier protein (ACP) is an essential cofactor for plant fatty acid synthesis. Three isoforms occur in barley seedling leaves. The genes Acl1 and Acl3 coding for the predominant ACP I and the minor ACP III, resp., have been cloned and characterized as has a full-length cDNA for ACP III. Both genes, extending over more than 2.5 kb, have a conserved mosaic structure of 4 exons and 3 introns which result in mRNAs of .apprx.900 bases. **Alignment** of the DNA **sequences** demonstrates that homol. is restricted to the 2 **exons** coding for the mature protein whereas the remaining segments of the genes including the transit peptide-coding domains lack homol. Southern blot

analyses demonstrate that Acl1 and Acl3 represent single copy genes located on chromosomes 7 and 1, resp. Primer extension analyses identified multiple transcription start sites in both genes. The promoter regions are remarkably different; that of Acl3 resembles those for mammalian housekeeping genes in having a high G + C content plus three copies of an RNA polymerase II recognition GC element and in lacking correctly positioned TATA boxes. These features are in accordance with the hypothesis that Acl1 is specifically expressed in leaf tissue whereas Acl3 is a constitutively expressed gene.

L6 ANSWER 48 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1992:404765 CAPLUS Full-text
DN 117:4765
TI Exon-skipping is responsible for the 9 amino acid residue deletion
occurring near the N-terminal of human β-casein
AU Martin, Patrice; Leroux, Christine
CS Lab. Genet. Biochim., Inst. Natl. Rech. Agron., Jouy-en-Josas, 78352, Fr.
SO Biochemical and Biophysical Research Communications (1992), 183(2), 750-7
CODEN: BBRCA9; ISSN: 0006-291X
DT Journal
LA English
AB Interspecies comparison and **alignment** of the β-casein N-terminal **sequence**, taking into account its **exon** modular splitting derived from the known structural organization of the relevant genes, has revealed that a 9 amino acid residue sequence, corresponding to that encoded by the third exon of the other species genes, is lacking in human β-casein. Using the polymerase chain reaction technique, the authors amplified a human genomic 1-kb fragment, spanning from exon 2 to exon 4, which was subsequently cloned and sequenced. One hundred base pairs (bp) upstream from exon 4 and 737 bp downstream of exon 2, a 27-kb virtual exon 3 sequence, probably skipped during the course of pre-mRNA splicing, was identified. The possibility that this out-splicing event might be due to the weak strength of the 3' acceptor site and/or to the secondary structure sequestering of the branch site sequence is discussed.

L6 ANSWER 49 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1992:188903 CAPLUS Full-text
DN 116:188903
TI Structure of the gene for human coagulation factor V
AU Cripe, Larry D.; Moore, Karen D.; Kane, William H.
CS Med. Cent., Duke Univ., Durham, NC, 27710, USA
SO Biochemistry (1992), 31(15), 3777-85
CODEN: BICHAW; ISSN: 0006-2960
DT Journal
LA English
AB Activated factor V (Va) serves as an essential protein cofactor for the conversion of prothrombin to thrombin by factor Xa. Anal. of the factor V cDNA indicates that the protein contains several types of internal repeats with the following domain structure: A1-A2-B-A3-C1-C2. In this report the isolation and characterization of genomic DNA coding for human factor V is described. The factor V gene contains 25 exons which range in size from 72 to 2820 bp. The structure of the gene for factor V is similar to the previously characterized gene for factor VIII. Based on the **aligned** amino acid **sequences** of the 2 proteins, 21 of the 24 intron-**exon** boundaries in the factor V gene occur at the same location as in the factor VIII gene. In both genes, the junctions of the A1-A2 and A2-A3 domains are each encoded by a single exon. In contrast, the boundaries between domains A3-C1 and C1-C2 occur at intron-exon boundaries, which is consistent with evolution through domain duplication and exon shuffling. The connecting region or B domain of factor V is encoded by a single large exon of 2820 bp. The corresponding exon of the factor VIII gene contains 3106 bp. The 5' and 3' ends of both of these exons encode sequences homologous to the carboxyl-terminal end of domain A2 and the amino-terminal end of domain A3 in ceruloplasmin. There is otherwise no homol. between the B domain exons. These data provide further insight into the evolutionary relationships within this family of related plasma proteins and provide a basis from which to begin the investigation of the cellular regulation of factor V biosynthesis and characterization of mol. defects in congenital factor V deficiency.

L6 ANSWER 50 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1992:122234 CAPLUS Full-text
DN 116:122234
TI Genomic DNA sequence of the cystic fibrosis transmembrane conductance
regulator (CFTR) gene
AU Zielenski, Julian; Rozmahel, Richard; Bozon, Dominique; Kerem, Bat Sheva;
Grzelczak, Zbyszko; Riordan, John R.; Rommens, Johanna; Tsui, Lap Chee
CS Dep. Genet., Hosp. Sick Child., Toronto, ON, M5G 1X8, Can.
SO Genomics (1991), 10(1), 214-28
CODEN: GNMCEP; ISSN: 0888-7543
DT Journal
LA English
AB The gene responsible for cystic fibrosis, the most common severe autosomal recessive disorder, is located on the long arm of human chromosome 7, region q31-q32. The gene has recently been identified and shown to be approx. 250 kb in size. To understand the structure and to provide the basis for a systematic anal. of the disease-causing mutations in the gene, genomic DNA clones spanning different regions of the previously reported cDNA were isolated and used to determine the coding regions and sequences of intron/exon boundaries. Total of 22,708 bp of sequence, accounting for .apprx.10% of the entire gene, was obtained. Alignment of the genomic DNA sequence with the cDNA sequence showed perfect colinearity between the two and a total of 27 exons, each flanked by consensus splice signals. A number of repetitive elements, including the Alu and Kpn families and simple repeats, such as (GT)17, (GATT)7, and (TA)14, were detected in close vicinity of some of the intron/exon boundaries. At least three of the simple repeats were found to be polymorphic in the population. Although an internal amino acid sequence homol. could be detected between the two halves of the predicted polypeptide, especially in the regions of the two putative nucleotide-binding folds (NBF1 and NBF2), the lack of **alignment** of the nucleotide **sequence** as well as the different positions of the **exon**/intron boundaries does not seem to support the hypothesis of a recent gene duplication event. To facilitate detection of mutations by direct sequence anal. of genomic DNA, 28 sets of oligonucleotide primers were designed and tested for their ability to amplify individual exons and the immediately flanking sequences in the introns.

L6 ANSWER 51 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1992:81719 CAPLUS Full-text
DN 116:81719

TI Evolution of the rabbit immunoglobulin κ chain genes
AU Ayadi, Hammadi; Marche, Patrice N.; Cazenave, Pierre Andre
CS Dep. Immunol., Inst. Pasteur, Paris, F-75742, Fr.
SO Immunogenetics (1991), 34(3), 201-7
CODEN: IMNGBK; ISSN: 0093-7711
DT Journal
LA English
AB The organization and the structure of rabbit κ chain genes encoding b allotypes were analyzed in wild rabbits. The κ1 gene of the b95 allotype was cloned and its structure determined The J region is composed of 5 segments but only J2 appears to be functional and is identical to the J2 segment of the b4 allotype. The J region is highly conserved among the various b allotypes, whereas the constant region exon displays a high level of differences when compared with other allotypes. The b95 J region is closer to that of b4var and the constant region to b5 allotype constant region. **Alignment** of nucleotide **sequences** revealed that the constant region **exon** displays segmental similarities with b4 and bas constant regions. The mosaic structure of b95 allotype gene indicates that complex allotypes of κ1 genes may result from genetic exchanges or gene conversion between the different κ genes.

L6 ANSWER 52 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1991:649227 CAPLUS Full-text
DN 115:249227
TI Genomic organization and sequence of a rat class I MHC gene that is an
apparent pseudogene
AU Kryspin-Sorensen, Ilona; Johansen, Teit; Kastern, William
CS Coll. Med., Univ. Florida, Gainesville, FL, 32610, USA
SO Immunogenetics (1991), 33(3), 213-15
CODEN: IMNGBK; ISSN: 0093-7711
DT Journal
LA English
AB A 7-kb BamHI genomic fragment which was missing from some inbred rat strains including the spontaneously diabetic BB/Wor (RT1u) rat and diabetes-prone lines and present in the diabetes-resistant lines was cloned. The complete gene sequence of clone RT1.A-4 which contains the fragment is reported. The gene is **aligned** on 8 **exons** similar to the mouse class I MHC genes **sequenced** thus far. The intron-exon junctions are similar to those of the mouse class I genes. Likewise, there is a high degree of similarity both at the nucleotide and amino acid levels with other typical class I genes. The most notable feature of RT1.A-4 is that a termination colon was detected within the second exon in a region that would code for the α1 domain of the protein (nucleotides 491-493). With the exception of this codon, sequence similarity with other class I MHC genes is maintained throughout this gene. No other termination codons are apparent apart from the typical one in the eighth exon (nucleotides 3206-3208, and there is every indication that this gene would have encoded a typical class I MHC antigen were it not for the premature termination codon in exon 2. The sequence includes 150 nucleotides upstream from the putative initiation codon (nucleotides 154-156) and 1000 nucleotides downstream from the putative polyadenylation signal (nucleotides 3646-3651).

L6 ANSWER 53 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1991:222477 CAPLUS Full-text
DN 114:222477
TI Structure of the human genomic region homologous to the bovine
prochymosin-encoding gene
AU Ord, T.; Kolmer, M.; Villems, R.; Saarma, M.
CS Inst. Chem. Phys. Biophys., Tallinn, USSR
SO Gene (1990), 91(2), 241-6
CODEN: GENED6; ISSN: 0378-1119
DT Journal
LA English
AB Two human genomic libraries were probed with bovine prochymosin (bPC) cDNA. Recombinant clones covering a genomic region homologous to the entire coding region and flanking sequences of the bPC gene were isolated. Human sequences homologous to exons of the bPC gene are distributed in a DNA fragment of 10 kb. **Alignment** of the human **sequences** and the **exons** of bPC reveals that the human '**exons**' 1-3, 5, and 7-9 have sizes identical to the corresponding bovine exons, but a nucleotide (nt) has been deleted in the human exon 4 and two nt in the human exon 6. The aligned human sequence and the coding part of the bPC gene share 82% nt homol., the value ranging, in sep. exons, from 76 (exon 1) to 84% (exons 5 and 6). The 150 bp of 5'-flanking sequence of the human gene has 75% homol. to the corresponding region of the bPC gene and contains a TATA-box in a similar position. A 1-nt deletion in the human exon 4 would shift the translational reading frame of a putative human PC mRNA relative to bPC mRNA, and result in an in-phase terminator spanning codons 163 and 164 in bPC mRNA. Another terminator in-phase with the amino-acid sequence encoded by the bPC gene occurs in the human exon 5 and the second frameshift mutation in exon 6. Thus, the nt sequence anal. of the human genomic region has revealed the presence of mutations that have rendered it unable to produce a full-length protein homologous to bPC and, therefore, this gene is considered as a human prochymosin pseudogene (hPCψ). Blot-hybridization anal. of human genomic DNA indicates that hPCψ is a single gene in the human genome.

L6 ANSWER 54 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1991:179447 CAPLUS Full-text
DN 114:179447
TI Gene segments encoding membrane domains of the human immunoglobulin gamma
3 and alpha chains
AU Bensmana, Mylene; Lefranc, Marie Paule
CS Lab. Immunogenet. Mol., Univ. Montpellier II, Montpellier, 34095, Fr.
SO Immunogenetics (1990), 32(5), 321-30
CODEN: IMNGBK; ISSN: 0093-7711
DT Journal
LA English
AB The c-terminal region of the heavy chains, according to its hydrophilic or hydrophobic properties, dets. whether the Ig will be secreted or membrane-bound. The nucleotide sequences of the human IGHG3, IGHA1, and IGHA2 membrane exons isolated from genomic DNA libraries were determined The IGHG3 M1 and M2 exons are separated by a long intron of 2.1 kilobases (kb) containing an highly repeated motif of 34 base pairs (bp). The IGHA1 and IGHA2 genes, like the mouse Igh-A gene, have a single exon encoding the extracellular, transmembrane, and cytoplasmic regions. For each class of Igs, the **sequences** of membrane **exons** are highly conserved between human and mouse, but no **alignment** is possible for the flanking regions. In contrast, for a same species, the sequences of the heavy chain membrane exons differ from one class to another. While the hydrophobic profile of the membrane core is well conserved, the cytoplasmic region differs in length and in composition None of the intracellular

domains presents the sequence implied in signal transduction, implying that membrane Igs need other proteins, which probably interact with the constant or membrane domain, to transmit signals leading to B-cell activation.

L6 ANSWER 55 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1990:472122 CAPLUS Full-text
DN 113:72122
TI Structural organization of the gene encoding apolipoprotein A-II in an
amyloidotic strain of senescence-accelerated mouse
AU Yonezu, Tomonori; Toda, Masaaki; Yamagishi, Hideo; Higuchi, Keiichi;
Takeda, Toshio
CS Gerontol. Nutr. Div., Meiji Inst. Health Sci., Odawara, 250, Japan
SO Gene (1989), 84(1), 187-91
CODEN: GENED6; ISSN: 0378-1119
DT Journal
LA English
AB An inherited polymorphism occurring in the murine apolipoprotein A-II (ApoA-II) transcript seems to be related to the senile amyloidosis which occurs in accelerated-senescence-prone mice (SAM-P). Such being the case, the entire nucleotide (nt) sequence of the apoA-II gene was determined The length of the gene is about 1.3 kb. It is interrupted by 3 introns; the 4 **exons align** perfectly with the previously **sequenced** elements of an apoA-II cDNA. Two nt substitutions [Pro-5(CCA) → Gln(CAG)] in the SAM-P genome were identified in the third exon, hence, a restriction fragment length polymorphism was used to detect the apoA-II mol. type. Several possible regulatory signals were identified (i) in the 5'-flanking region, including CAAT and TATA boxes, the viral enhancer-like sequence, and the consensus sequences of estrogen response element, and (ii) in the 3'-flanking region, including sequences conserved in the immunoglobulin enhancer, glucocorticoid and estrogen response elements, and a B1 repetitive sequence.

L6 ANSWER 56 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1990:453404 CAPLUS Full-text
DN 113:53404
TI The Balbiani ring 3 gene in Chironomus tentans has a diverged repetitive
structure split by many introns
AU Paulsson, Gabrielle; Lendahl, Urban; Galli, Joakim; Ericsson, Christer;
Wieslander, Lars
CS Dep. Mol. Genet., Karolinska Inst., Stockholm, S-104 01, Swed.
SO Journal of Molecular Biology (1990), 211(2), 331-49
CODEN: JMOBAK; ISSN: 0022-2836
DT Journal
LA English
AB A set of approx. 15 secretory proteins is synthesized by the salivary gland cells in the midge C. tetans. These proteins are secreted but do not form insol. fibers until they are transported out of the gland lumen. A Balbiani ring (BR) gene family consisting of four genes (BR1, BR2.1, BR2.2 and BR6) have previously been shown to encode 4 of these proteins, sp-I a to d, with relative mol. wts. of 1 + 106. Each BR gene contains an uninterrupted block in which about 100 repeats are tandemly arranged. The repeats are virtually identical and efficient homogenization mechanisms must operate within each block. The BR3 gene, which according to structural similarities may belong to the BR gene family, but at the same time exhibits a strikingly different structure is described here. The gene encodes a 10.9 kb transcript that contains 38 introns and is spliced into a 5.5 kb mRNA. The mRNA is translated into a cysteine-rich 185 kDa major component of the gland secretion. The coding sequence in the gene is built from diverged repeats in which mainly the cysteine codons are preserved and the sequence is split by the introns into 17 to 678-bp long exons. The introns are located at defined positions in relation to the repeat structure. In sharp contrast to the uninterrupted array of identical repeats in the BR1-BR6 genes, the repeats in the BR3 gene are not efficiently homogenized and have diverged extensively from each other. The splitting of the repeat structure into variable sized **exons** may prevent homogenizations dependent on unequal **aligning** of homologous **sequences.**

L6 ANSWER 57 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1990:152798 CAPLUS Full-text
DN 112:152798
TI Human CYP1A2: sequence, gene structure, comparison with the mouse and rat
orthologous gene, and differences in liver 1A2 mRNA expression
AU Ikeya, Kiyoko; Jaiswal, Anil K.; Owens, Roland A.; Jones, John E.; Nebert,
Daniel W.; Kimura, Shioko
CS Lab. Dev. Pharmacol., Natl. Inst. Child Health Hum. Dev., Bethesda, MD,
20892, USA
SO Molecular Endocrinology (1989), 3(9), 1399-408
CODEN: MOENEN; ISSN: 0888-8809
DT Journal
LA English
AB The human CYP1A2 (cytochrome P3450) gene and 1906 basepairs (bp) of the 5' flanking and 113 bp of the 3' flanking regions were sequenced. The gene spans almost 7.8 kilobases, comprising 7 exons and 6 introns. The transcriptional start site was determined by both primer extension and S1 mapping. Including the first noncoding exon of 55 bp, the entire mRNA is 3121 bp in length, and the open reading frame, starting with nucleotide 10 of exon 2, encodes 515 amino acids (mol. weight = 58,294). Between the human CYP1A2 and CYP1A1 (cytochrome P1450) genes, exons 2, 4, 6, and especially 5 are strikingly conserved in both nucleotide similarity and total number of bases. **Alignment** of the upstream **sequences** and **exon** 1 of human CYP1A2 with that of mouse or rat CYP1A2 revealed 2 possibly significant regions of similarity: 1) 68% in the .apprx.150 bases immediately 5' from the mRNA cap site and 2) 80% identity between the human -841 to -758 segment and the mouse -1529 to -1439 segment. The canonical 5-bp box (CACGC), found upstream of all mammalian CYP1A1 genes to date and believed to interact with the inducer-aromatic hydrocarbon receptor complex, was not found on either strand in the 1906 bp of the 5' flanking region of human CYP1A2. In contrast, **alignment** of the upstream **sequences**, **exon** 1, and intron 1 of human CYP1A1 with that of mouse or rat CYP1A1 revealed large, highly conserved regions. Conserved regions were found in intron 1 of the human, mouse, and rat CYP1A2 gene. These data suggest that the regulatory elements controlling the CYP1A2 gene might differ in location from those controlling the CYP1A1 gene. Among 12 human liver samples, striking differences (>15-fold) in the 3.3-kilobase 1A2 mRNA levels were seen. This result may reflect significant genetic differences in constitutive and/or inducible CYP1A2 gene expression that could play an important role in individual risk of environmental toxicity or cancer.

somatic-cell hybrid containing only a human isochromosome 12p in a mouse background.

L6 ANSWER 58 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1990:71423 CAPLUS Full-text
DN 112:71423
TI Partial nucleotide sequence of a bovine major histocompatibility class II
DRβ-like gene
AU Muggli-Cockett, N. E.; Stone, R. T.
CS U. S. Meat Anim. Res. Cent., ARS, Clay Center, NE, USA
SO Animal Genetics (1989), 20(4), 361-9
CODEN: ANGEE3; ISSN: 0268-9146
DT Journal
LA English
AB A genomic clone contg. a bovine DRβ-like gene, BoDRβII, was isolated from a bovine genomic library and characterized by restriction enzyme mapping and nucleotide sequencing of exon regions. **Alignment** of this **sequence** with the human DRβ cDNA **sequence** allowed identification of **exon**/intron boundaries. The clone contains a 13.3-kilobase (kb) insert, and includes 1.3 kb 5' of β1 exon and 6.7 kb 3' of the transmembrane (TM) exon. Open reading frames were present in the BoDRβ exons sequenced. Nucleotide identities of the bovine β1, β2, and TM exons with the corresponding human DRβ exons were 73, 91, and 83%, resp. Nucleotide identities of these exons with those of a previously described bovine DRβ-like pseudogene, BoDRβI, were 69, 95, and 81%, resp. Although a limited amount of sequence data was obtained for the intron regions, a 71% identity was found within a 514-nucleotide region immediately 3' to the β2 exons in BoDRβI and BoDRβII. A series of GT residues followed by a longer series of GA residues began ≈35 nucleotides 3' of the β1 exon in both BoDRβI and BoDRβII.

L6 ANSWER 59 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1989:610017 CAPLUS Full-text
DN 111:210017
TI A cluster of α2-macroglobulin-related genes (α2M) on human chromosome 12p: cloning of the pregnancy-zone protein gene and an
α2M pseudogene
AU Devriendt, Koen; Zhang, Ji; Van Leuven, Fred; Van den Berghe, Herman;
Cassiman, Jean Jacques; Marynen, Peter
CS Cent. Hum. Genet., Univ. Leuven, Louvain, B-3000, Belg.
SO Gene (1989), 81(2), 325-34
CODEN: GENED6; ISSN: 0378-1119
DT Journal
LA English
AB The characterization of two α2-macroglobulin(α2M)-related genomic clones, isolated from two human genomic libraries by use of α2M cDNA as a probe, is reported. Sequence comparison of the clone EPZP6 with the human α2M cDNA revealed the presence of five exons with the proper splice signals. **Alignment** of the corresponding amino acid (aa) **sequence** of these **exons** with the published partial pregnancy-zone protein (PZP) aa **sequence** showed a perfect match, thereby identifying EPZP6 as a PZP genomic clone. The clone MPAM16 showed a considerable degree of sequence conservation when compared to the human α2M cDNA sequence, and several putative exons were identified. However, a frameshift mutation leading to a premature stop codon was found in the coding sequence, classifying this gene as an α2M pseudogene. Human α2M, PZP and the related pseudogene were mapped to the human chromosome 12p12-13, with the help of gene-specific probes and in situ hybridization. This result was confirmed in Southern-blot expts. with DNA from a human-Ltk- mouse

L6 ANSWER 60 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1989:129566 CAPLUS Full-text
DN 110:129566
TI Rudimentary phosvitin domains in a minor chicken vitellogenin gene
AU Byrne, B. Marion; De Jong, Harry; Fouchier, Ronaldus A. M.; Williams,
David L.; Gruber, Max; Ab, Geert
CS Biochem. Lab., Groningen Univ., Groningen, 9747 AG, Neth.
SO Biochemistry (1989), 28(6), 2572-7
CODEN: BICHAW; ISSN: 0006-2960
DT Journal
LA English
AB The nucleotide sequence and the derived amino acid sequence of the phosphoprotein-encoding region of the chicken vitellogenin III gene were determined The **sequence** of this minor vitellogenin could be **aligned** with **exon** 22 up to **exon** 27 of the previously **sequenced** major vitellogenin II gene. The exon 23 and 25 sequences are rich in serine codons (26% and 41%, resp.), and this region encodes at least one of the small egg yolk phosphoproteins. The major egg yolk phosphoprotein, phosvitin, is encoded by the analogous region in vitellogenin II. Comparison of the vitellogenin II and vitellogenin III sequences shows a great reduction in the size of the putative exon 23 of the latter (321 base pairs as opposed to 690). The number of serine codons is also drastically reduced from 124 in exon 23 of the vitellogenin II gene to 28 in vitellogenin III. The grouping of synonymous serine codons, as has hitherto been observed in sequenced vitellogenin phosphoproteins, has been maintained in vitellogenin III. A putative asparagine-linked N-glycosylation site which was conserved in the chicken vitellogenin II and the Xenopus laevis vitellogenin A2 gene, at the beginning of exon 23, is also present in vitellogenin III. The two chicken vitellogenins show a low conservation in the phosphoprotein- encoding region (average 33%, at the protein level) compared to that in the peripheral sequences (58% identity), which indicates that it is a rapidly evolving domain of the vertebrate vitellogenin gene.

L6 ANSWER 61 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1987:569782 CAPLUS Full-text
DN 107:169782
TI Nucleotide sequence and organization of the human S-protein gene:
repeating peptide motifs in the "pexin" family and a model for their
evolution
AU Jenne, Dieter; Stanley, Keith K.
CS Inst. Med. Microbiol., Justus-Liebig-Univ., Giessen, 6300, Fed. Rep. Ger.
SO Biochemistry (1987), 26(21), 6735-42
CODEN: BICHAW; ISSN: 0006-2960
DT Journal
LA English
AB The S-protein/vitronectin gene was isolated from a human genomic DNA library, and its sequence of .apprx.5.3 kilobases, including the adjacent 5'- and 3'-flanking regions, was established. **Alignment** of the genomic DNA nucleotide sequence and the cDNA **sequence** indicated that the gene consisted of eight **exons** and seven introns. The intron positions in the S-protein gene and their phase type were compared to those in the hemopexin gene which shares amino acid sequence homologies with transin and

the S-protein. Three introns were found at an equivalent position; two other introns are very close to these positions and are interpreted as cases of intron sliding. Introns 3-7 occur at a conserved glycine residue within repeating peptide segments, whereas introns 1 and 2 are at the boundaries of the somatomedin B domain of S-protein. The anal. of the exon structure in relation to repeating peptide motifs within the S-protein strongly suggests that it contains only seven repeats, one less than the hemopexin mol. A very similar repeat pattern like that in hemopexin is present also in two other related proteins, transin and interstitial collagenase. An evolutionary model for the generation of the repeat pattern in the S-protein and the other members of this novel "pexin" gene family is proposed, and the sequence modifications for some of the repeats during divergent evolution are discussed in relation to known unique functional properties of hemopexin and S-protein.

L6 ANSWER 62 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1987:510238 CAPLUS Full-text
DN 107:110238
TI Isolation, structure and expression of mammalian genes for histidyl-tRNA synthetase
AU Tsui, Florence W. L.; Siminovitch, L.
CS Mount Sinai Hosp. Res. Inst., Toronto, ON, M5G 1X5, Can.
SO Nucleic Acids Research (1987), 15(8), 3349-67
CODEN: NARHAD; ISSN: 0305-1048
DT Journal
LA English
AB A full-length cDNA clone that codes for human histidyl-tRNA synthetase (HRS) and cDNA clones that span the full-length transcript of hamster HRS were isolated. The full-length human HRS cDNA was expressed after transfection into Cos 1 cells and a CHO ts mutant defective in the gene for HRS. The complete nucleotide sequence of the hamster and human gene were obtained, and extensive homologies were observed in three regions on comparing these sequences between themselves and with the sequence of HRS derived from yeast. These results provide unequivocal evidence that the hamster and human genes for HRS were cloned. Three overlapping phage recombinants containing the complete hamster chromosomal gene for HRS were also isolated. The genomic HRS is divided into 13 exons. The precise locations of each of the 5' and 3' **exon**-intron boundaries were defined by **sequencing** the appropriate regions of the cloned genomic DNA and **aligning** them with the sequence of HRS cDNAs. These studies provide the basis for future structural and functional anal. of the gene for HRS. In particular, it will be of interest to examine if different exons of HRS correlate to different domains of the HRS polypeptide.

L6 ANSWER 63 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1986:492342 CAPLUS Full-text
DN 105:92342
TI Structure and expression of the mouse prealbumin gene
AU Wakasugi, Shoji; Maeda, Shuichiro; Shimada, Kazunori
CS Med. Sch., Kumamoto Univ., Kumamoto, 860, Japan
SO Journal of Biochemistry (Tokyo, Japan) (1986), 100(1), 49-58
CODEN: JOBIAO; ISSN: 0021-924X
DT Journal
LA English
AB A genomic DNA fragment was cloned which covers the entire sequence of the mouse prealbumin gene, and the structure was studied. The coding regions are separated into 4 exons by 3 introns; and these nos., the sizes of the exons, and the relative sites of the exon-intron junctions are all in complete agreement with those determined for the human gene. The **sequences** of 4 **exons** can be **aligned** perfectly with that of the previously determined mouse prealbumin cDNA. In addition to the exon regions, 2 highly conserved DNA regions were found between the mouse and human prealbumin genes, one in the 5'-flanking region of the gene and the other in the 3' end region of the first intron. These DNA regions contain several consensus glucocorticoid receptor-binding site sequences, and the latter also contains an enhancer sequence present in the Ig kappa-chain joining-constant intron. RNA hybridizing to the mouse prealbumin cDNA was detected in the exts. from liver, brain, and kidney but was not detected in testes, spleen, or heart. Little change was caused in the level of prealbumin mRNA in the liver by administration of dexamethasone to mice.

L6 ANSWER 64 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1986:436626 CAPLUS Full-text
DN 105:36626
TI Sequence, topography and protein coding potential of mouse int-2: a putative oncogene activated by mouse mammary tumour virus
AU Moore, R.; Casey, G.; Brookes, S.; Dixon, M.; Peters, G.; Dickson, C.
CS Imp. Cancer Res. Fund Lab., London, WC2A 3PX, UK
SO EMBO Journal (1986), 5(5), 919-24
CODEN: EMJODG; ISSN: 0261-4189
DT Journal
LA English
AB A major proportion of carcinomas induced by mouse mammary tumor virus (MMTV) show evidence for proviral activation of a cellular gene, int-2, on chromosome 7. The sequence of base pairs of DNA spanning the transcription unit of int-2 was determined and compared with that of a series of int-2-specific cDNA clones derived from mammary tumor RNA. The predicted positions of intron-**exon** boundaries, established by **alignment** of cDNA and chromosomal DNA **sequences**, indicate that the gene comprises ≥3 **exons.** An open reading frame capable of encoding a protein of 245 amino acids with an estimated mol. weight of 27 kilodaltons, is flanked by substantial noncoding segments at both 5' and 3' ends. Comparison of the chromosomal DNA sequence and the predicted amino acid sequence with available data bases has revealed no homol. to other known genes. These results are discussed in relation to the status of int-2 as a candidate protooncogene.

L6 ANSWER 65 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1986:162855 CAPLUS Full-text
DN 104:162855
TI Immunological identification of rat tissue kallikrein cDNA and characterization of the kallikrein gene family
AU Gerald, William L.; Chao, Julie; Chao, Lee
CS Dep. Biochem., Med. Univ. South Carolina, Charleston, SC, 29425, USA
SO Biochimica et Biophysica Acta (1986), 866(1), 1-14
CODEN: BBACAQ; ISSN: 0006-3002
DT Journal
LA English
AB A tissue kallikrein [9001-01-8] cDNA was identified by direct immunol. screening with affinity-purified anti-rat tissue kallikrein antibody from a rat submandibular cDNA library constructed with the expression vector pUC8. Sequence anal. of the kallikrein cDNA revealed an encoded protein 97% homologous to the partial amino acid

sequence of rat submandibular kallikrein. This cDNA was used to hybrid-select kallikrein-specific RNA from submandibular gland. Translation of the hybrid-selected RNA in a cell-free assay system resulted in the production of a 37-kilodalton peptide representing the preproenzyme. In addition, hybrid-selection of RNA under less stringent conditions showed cross-hybridization with other submandibular gland mRNA species. In correlation with these results, anal. of rat genomic DNA showed extensive hybridization, suggesting a family of closely related kallikrein-like genes. Consequently, a Charon 4A rat genomic library was screened for kallikrein genes by hybridization with rat tissue kallikrein cDNA. Thirty-four clones were isolated and found to be highly homologous by hybridization and restriction enzymes analyses. Fourteen unique clones were identified by restriction enzyme site polymorphisms within DNA segments which hybridized to the kallikrein cDNA probe and it was estimated that 17 different kallikrein-like genes are present in the rat. Sequence and structural anal. of one of the genomic clones revealed a gene structure similar to that of other serine proteinases. Comparison of the partially sequenced **exon** regions of the gene with the **sequence** of rat tissue kallikrein cDNA reveals 89% identity when **aligned** for the greatest homol. However, the genomic sequence predicts termination codons in all 3 translational reading frames, implying that this gene is nonfunctional, i.e., a pseudogene. Comparison of the rat genomic sequence to a kallikrein-like gene from the mouse reveals extensive preservation of exons, less identity within introns, and no significant homol. between extragenic regions.

L6 ANSWER 66 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1985:143999 CAPLUS Full-text
DN 102:143999
TI Organization and structure of the mouse interleukin-2 gene
AU Fuse, Akira; Fujita, Takashi; Yasumitsu, Hidetaro; Kashima, Nobukazu;
Hasegawa, Katsushige; Taniguchi, Tadatsugu
CS Cancer Inst., Jpn. Found. Cancer Res., Tokyo, 170, Japan
SO Nucleic Acids Research (1984), 12(24), 9323-31
CODEN: NARHAD; ISSN: 0305-1048
DT Journal
LA English
AB A chromosomal DNA segment was cloned which covers the entire sequence for the murine interleukin-2 (IL-2) gene, and its structure was analyzed. The coding regions are separated into 4 blocks by 3 introns, each of which is located similarly to the corresponding human gene. The **exon sequences** can be **aligned** perfectly with the previously cloned cDNA **sequence.** Of particular interest is the presence of sequences within the 5'-flanking region which are highly conserved between mouse and man. The conserved region which spans >400 base pairs may play a role in the regulation of IL-2 gene expression.

L6 ANSWER 67 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1984:80623 CAPLUS Full-text
DN 100:80623
TI Structure of the human interleukin 2 gene
AU Fujita, Takashi; Takaoka, Chikako; Matsui, Hiroshi; Taniguchi, Tadatsugu
CS Cancer Inst., Jap. Found. Cancer Res., Tokyo, 170, Japan
SO Proceedings of the National Academy of Sciences of the United States of
America (1983), 80(24), 7437-41
CODEN: PNASA6; ISSN: 0027-8424
DT Journal
LA English
AB Two species of EcoRI-cleaved DNA segments that together cover the entire sequence for the human interleukin 2 gene were cloned and the nucleotide sequence of the gene and its flanking regions were determined The gene contains 3 introns and the **exon sequences** can be **aligned** with the previously reported cDNA **sequence** almost perfectly except for a few nucleotides in the 3' nontranslated region. The promoter region contains a prototype TATA sequence as well as a notable palindromic sequence. Particularly interesting is the presence of sequences in this region that are homologous to the promoter region of the human interferon-γ gene. In addition, a sequence that closely resembles the core sequence for the viral enhancer elements has been found in the 2nd intron. Such sequences may play a role in the expression of the interleukin 2 gene in lectin- or antigen-stimulated T lymphocytes.

L6 ANSWER 68 OF 68 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1983:535697 CAPLUS Full-text
DN 99:135697
TI Close relationship between certain nuclear and mitochondrial introns.
Implications for the mechanism of RNA splicing
AU Waring, R. B.; Scazzocchio, C.; Brown, T. A.; Davies, R. W.
CS Inst. Sci. Technol., Univ. Manchester, Manchester, M60 1QD, UK
SO Journal of Molecular Biology (1983), 167(3), 595-605
CODEN: JMOBAK; ISSN: 0022-2836
DT Journal
LA English
AB The 1st indication of a direct relation between a nuclear and a mitochondrial splicing system is presented. The intron in the precursor of the large, nuclearly coded rRNA of 2 species of Tetrahymena possesses all the features of a class of fungal mitochondrial introns. Sequences conserved in mitochondrial introns of different fungal species are also found in the same order in these Tetrahymena nuclear introns, and the intron RNA can be folded to form a secondary structure similar to that previously proposed for mitochondrial introns. This core secondary structure brings the ends of the intron together. Furthermore, the 1st intron in the precursor of the large, nuclearly coded rRNA of Physarum polycephalum also has the characteristic conserved sequences and core RNA secondary structure. The limited sequence data available suggest that the intron in the large rRNA of chloroplasts in Chlamydomonas reinhardtii also resembles the mitochondrial introns. Tetrahymena Large nuclear rRNA introns also have an internal sequence that can act as an adapter by pairing with upstream and downstream **exon sequences** adjacent to the splice junctions to **align** precisely the splice junctions. These nuclear introns therefore fit a previously proposed model of the role of intron RNA in the splicing process, which suggests that the mechanisms of splicing may be very similar in these apparently diverse systems. Thus, the RNA secondary structures, for which there is good evidence in the case of mitochondrial introns, may form the basis of active site structure and precise alignment in splicing and cyclization of the Tetrahymena intron ribozyme.

=> log y

COST IN U.S. DOLLARS	SINCE FILE ENTRY	SESSION	TOTAL
FULL ESTIMATED COST		197.65	197.86

DISCOUNT AMOUNTS (FOR QUALIFYING ACCOUNTS) SINCE FILE TOTAL

	ENTRY	SESSION
CA SUBSCRIBER PRICE	-47.12	-47.12

STN INTERNATIONAL LOGOFF AT 18:55:43 ON 12 MAY 2004

```
********** Welcome to STN International ********
**

 NEWS  1       Web Page URLs for STN Seminar Schedule - N.
America
 NEWS  2       "Ask CAS" for self-help around the clock
 NEWS  3 JAN 27  Source of Registration (SR) information in
REGISTRY updated
             and searchable
 NEWS  4 JAN 27  A new search aid, the Company Name Thesaurus,
available in
             CA/CAplus
 NEWS  5 FEB 05  German (DE) application and patent publication
number format
             changes
 NEWS  6 MAR 03  MEDLINE and LMEDLINE reloaded
 NEWS  7 MAR 03  MEDLINE file segment of TOXCENTER reloaded
 NEWS  8 MAR 03  FRANCEPAT now available on STN
 NEWS  9 MAR 29  Pharmaceutical Substances (PS) now available
on STN
 NEWS 10 MAR 29  WPIFV now available on STN
 NEWS 11 MAR 29  New monthly current-awareness alert (SDI)
frequency in RAPRA
 NEWS 12 APR 26  PROMT: New display field available
 NEWS 13 APR 26  IFIPAT/IFIUDB/IFICDB: New super search and
display field
             available
 NEWS 14 APR 26  LITALERT now available on STN
 NEWS 15 APR 27  NLDB: New search and display fields available
 NEWS 16 May 10  PROUSDDR now available on STN
 NEWS 17 May 19  PROUSDDR: One FREE connect hour, per
account, in both May
             and June 2004
 NEWS 18 May 12  EXTEND option available in structure searching
 NEWS 19 May 12  Polymer links for the POLYLINK command
completed in REGISTRY

 NEWS EXPRESS  MARCH 31 CURRENT WINDOWS VERSION IS
V7.00A, CURRENT
          MACINTOSH VERSION IS V6.0c(ENG) AND V6.0Jc(JP),
          AND CURRENT DISCOVER FILE IS DATED 26 APRIL 2004
 NEWS HOURS    STN Operating Hours Plus Help Desk Availability
 NEWS INTER    General Internet Information
 NEWS LOGIN    Welcome Banner and News Items
 NEWS PHONE    Direct Dial and Telecommunication Network Access
to STN
 NEWS WWW      CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on
that
specific topic.

  All use of STN is subject to the provisions of the STN Customer
  agreement.  Please note that this agreement limits use to scientific
  research.  Use for software development or design or
implementation
  of commercial gateways or other similar uses is prohibited and may
  result in loss of user privileges and other penalties.

  ********** STN Columbus **********
**

FILE 'HOME' ENTERED AT 19:10:30 ON 12 MAY 2004

=> file caplus
COST IN U.S. DOLLARS                      SINCE FILE     TOTAL
                                        ENTRY   SESSION
FULL ESTIMATED COST                        0.21      0.21

FILE 'CAPLUS' ENTERED AT 19:10:41 ON 12 MAY 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER
AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2004 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is
held by the publishers listed in the PUBLISHER (PB) field (available
for records published or updated in Chemical Abstracts after
December
26, 1996), unless otherwise indicated in the original publications.
The CA Lexicon is the copyrighted intellectual property of the
American Chemical Society and is provided to assist you in searching
databases on STN.  Any dissemination, distribution, copying, or
storing
of this information, without the prior written consent of CAS, is
strictly prohibited.

FILE COVERS 1907 - 12 May 2004  VOL 140 ISS 20
FILE LAST UPDATED: 11 May 2004  (20040511/ED)

  This file contains CAS Registry Numbers for easy and accurate
  substance identification.


=> s ((mrna or cdna or est) and alignment and genomic)/bi,ab
      237766 MRNA/BI
      203436 MRNA/AB
      163879 CDNA/BI
      126366 CDNA/AB
      124556 EST/BI
      120039 EST/AB
       51827 ALIGNMENT/BI
       42578 ALIGNMENT/AB
       84397 GENOMIC/BI
       76296 GENOMIC/AB
L1       619 ((MRNA OR CDNA OR EST) AND ALIGNMENT AND
GENOMIC)/BI,AB

=> s l1 and align?/bi,ab
       96740 ALIGN?/BI
       87143 ALIGN?/AB
L2       619 L1 AND ALIGN?/BI,AB

=> s l2 and algorithm?/bi,ab
       76518 ALGORITHM?/BI
       65008 ALGORITHM?/AB
L3        33 L2 AND ALGORITHM?/BI,AB

=> s l3 not 2004/py
      385703 2004/PY
L4        30 L3 NOT 2004/PY

=> s l4 not 2003/py
     1135324 2003/PY
L5        23 L4 NOT 2003/PY

=> s l5 not 2002/py
     1127076 2002/PY
L6        19 L5 NOT 2002/PY

=> s l6 not 2001/py
     1076644 2001/PY
L7        15 L6 NOT 2001/PY

=> d his

    (FILE 'HOME' ENTERED AT 19:10:30 ON 12 MAY 2004)

    FILE 'CAPLUS' ENTERED AT 19:10:41 ON 12 MAY 2004
```

L1 619 S ((MRNA OR CDNA OR EST) AND ALIGNMENT AND GENOMIC)/BI,AB
L2 619 S L1 AND ALIGN?/BI,AB
L3 33 S L2 AND ALGORITHM?/BI,AB
L4 30 S L3 NOT 2004/PY
L5 23 S L4 NOT 2003/PY
L6 19 S L5 NOT 2002/PY
L7 15 S L6 NOT 2001/PY

=> d l7 1-15 bib ab

L7 ANSWER 1 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:517690 CAPLUS Full-text
DN 134:26042
TI JESAM: CORBA software components to create and publish **EST alignments** and clusters
AU Parsons, J. D.; Rodriguez-Tome, P.
CS Wellcome Trust Genome Campus, EMBL-Outstation, The European Bioinformatics Institute (EBI), Cambridge, CB10 1SD, UK
SO Bioinformatics (2000), 16(4), 313-325
CODEN: BOINFP; ISSN: 1367-4803
PB Oxford University Press
DT Journal
LA English
AB Expressed Sequence Tags (ESTs) are cheap, easy and quick to obtain relative to full **genomic** sequencing and currently sample more eukaryotic genes than any other data source. They are particularly useful for developing Sequence Tag Sites (STSs for mapping), polymorphism discovery, disease gene hunting, mass spectrometer proteomics, and most ironically for finding genes and predicting gene structure after the great effort of **genomic** sequencing. However, ESTs have many problems and the public **EST** databases contain all the errors and high redundancy intrinsic to the submitted data so it is often found that derived database views, which reduce both errors and redundancy, are more effective starting points for research than the original raw submissions. Existing derived views such as **EST** cluster databases and consensus databases have never published supporting evidence or intermediary results leading to difficulties trusting, correcting, and customizing the final published database. These difficulties have lead many groups to wastefully repeat the complex intermediary work of others in order to offer slightly different final views. A better approach might be to discover the most expensive common calcns. used by all the approaches and then publish all intermediary results. Given a globally accessible database with a suitable component interface, like the JESAM software described in this paper, the creation of customized **EST** -derived databases could be achieved with min. effort. Databases of **EST** and full-length **mRNA** sequences for four model organisms have been self-compared by searching for overlaps consistent with contiguity. The sequence comparisons are performed in parallel using a PVM process farm and previous results are stored to allow incremental updates with minimal effort. The overlap databases have been published with CORBA interfaces to enable flexible global access as demonstrated by example Java applet browsers. Simple **cDNA** supercluster databases built as **alignment** database clients are themselves published via CORBA interfaces browsable with prototypical applets. A comparison with UniGene Mouse and Rat databases revealed undesirable features in both and the advantages of contrasting perspectives on complex data.
RE.CNT 54 THERE ARE 54 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 2 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:452087 CAPLUS Full-text
DN 134:217670
TI Optimal spliced **alignment** of homologous **cDNA** to a **genomic** DNA template
AU Usuka, Jonathan; Zhu, Wei; Brendel, Volker
CS Department of Chemistry, Stanford University, Stanford, CA, 94305, USA
SO Bioinformatics (2000), 16(3), 203-211
CODEN: BOINFP; ISSN: 1367-4803
PB Oxford University Press
DT Journal
LA English
AB Motivation: Supplementary **cDNA** or **EST** evidence is often decisive for discriminating between alternative gene predictions derived from computational sequence inspection by any of a number of requisite programs. Without addnl. exptl. effort, this approach must rely on the occurrence of cognate ESTs for the gene under consideration in available, generally incomplete, **EST** collections for the given species. In some cases, particular exon assignments can be supported by sequence matching even if the **cDNA** or **EST** is produced from non-cognate **genomic** DNA, including different loci of a gene family or homologous loci from different species. However, marginally significant sequence matching alone can also be misleading. We sought to develop an **algorithm** that would simultaneously score for predicted intrinsic splice site strength and sequence matching between the **genomic** DNA template and a related **cDNA** or **EST** . In this case, weakly predicted splice sites may be chosen for the optimal scoring spliced **alignment** on the basis of surrounding sequence matching. Strongly predicted splice sites will enter the optimal spliced **alignment** even without strong sequence matching. Results: We designed a novel **algorithm** that produces the optimal spliced **alignment** of a **genomic** DNA with a **cDNA** or **EST** based on scoring for both sequence matching and intrinsic splice site strength. By example, we demonstrate that this combined approach appears to improve gene prediction accuracy compared with current methods that rely only on either search by content and signal or on sequence similarity.
RE.CNT 19 THERE ARE 19 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 3 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:417853 CAPLUS Full-text
DN 134:173510
TI An approach to alternative gene transcripts discovery through assembly of fragmental **cDNA** sequences
AU Irie, Ryotaro; Masuho, Yasuhiko; Nagai, Keiichi
CS Helix Research Institute, Inc., Chiba, 292-0812, Japan
SO Frontiers Science Series (2000), 30(Currents in Computational Molecular Biology), 109-110
CODEN: FCFUEO; ISSN: 0915-8502
PB Universal Academy Press, Inc.
DT Journal
LA English
AB Genome sequencing efforts for human, which will completed in a few years, have opened the post-**genomic** age where the function of genes will need to be explored. It would be helpful or necessary to know the transcription multiplicity and alternative transcripts of each gene before one proceeds to the detailed study of the gene. To address this issue, we have developed a DNA sequence assembly program which is tailored for assembling

fragmental **cDNA** sequences to create all contigs (sets of consistently **aligned** fragments) each of which could correspond to a transcript. In this study we describe the major steps of our DNA sequence assembly **algorithm** (called MakeAllContigs), and applied MakeAllContigs to some UniGene clusters.

RE.CNT 4 THERE ARE 4 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 4 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 2000:28991 CAPLUS Full-text
DN 132:162007
TI Frequent alternative splicing of human genes
AU Mironov, Andrey A.; Fickett, James Wildon; Gelfand, Mikhail S.
CS NIIGenetika, State Center of Biotechnology, Moscow, 113545, Russia
SO Genome Research (1999), 9(12), 1288-1293
CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB Alternative splicing can produce variant proteins and expression patterns as different as the products of different genes, yet the prevalence of alternative splicing has not been quantified. Here the spliced **alignment algorithm** was used to make a first inventory of exon-intron structures of known human genes using **EST** contigs from the TIGR Human Gene Index. The results on any one gene may be incomplete and will require verification, yet the overall trends are significant. Evidence of alternative splicing was shown in 35% of genes and the majority of splicing events occurred in 5' untranslated regions, suggesting wide occurrence of alternative regulation. Most of the alternative splices of coding regions generated addnl. protein domains rather than alternating domains.

RE.CNT 14 THERE ARE 14 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 5 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1999:797780 CAPLUS Full-text
DN 132:132876
TI A general approach to single-nucleotide polymorphism discovery
AU Marth, Gabor T.; Korf, Ian; Yandell, Mark D.; Yeh, Raymond T.; Gu, Zhijie;
Zakeri, Hamideh; Stitziel, Nathan O.; Hillier, LaDeana; Kwok, Pui-Yan;
Gish, Warren R.
CS Department of Genetics and Genome Sequencing Center, Washington
University, St. Louis, MO, USA
SO Nature Genetics (1999), 23(4), 452-456
CODEN: NGENEC; ISSN: 1061-4036
PB Nature America
DT Journal
LA English
AB Single-nucleotide polymorphisms (SNPs) are the most abundant form of human genetic variation and a resource for mapping complex genetic traits. The large volume of data produced by high-throughput sequencing projects is a rich and largely untapped source of SNPs. We present here a unified approach to the discovery of variations in genetic sequence data of arbitrary DNA sources. We propose to use the rapidly emerging **genomic** sequence as a template on which to layer often unmapped, fragmentary sequence data and to use base quality values to discern true allelic variations from sequencing errors. By taking advantage of the **genomic** sequence we are able to use simpler yet more accurate methods for sequence organization: fragment clustering, paralogue identification and multiple **alignment.** We analyze these sequences with a novel, Bayesian inference engine, POLYBAYES, to calculate the probability that a given site is polymorphic. Rigorous treatment of base quality permits completely automated evaluation of the full length of all sequences, without limitations on **alignment** depth. We demonstrate this approach by accurate SNP predictions in human ESTs **aligned** to finished and working-draft quality **genomic** sequences, a data set representative of the typical challenges of sequence-based SNP discovery.

RE.CNT 24 THERE ARE 24 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 6 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1999:676311 CAPLUS Full-text
DN 132:31553
TI Comparative analysis of noncoding regions of 77 orthologous mouse and
human gene pairs
AU Jareborg, Niclas; Birney, Ewan; Durbin, Richard
CS The Sanger Centre, Cambridge, CB10 1SA, UK
SO Genome Research (1999), 9(9), 815-824
CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB A data set of 77 **genomic** mouse/human gene pairs has been compiled from the EMBL nucleotide database, and their corresponding features determined This set was used to analyze the degree of conservation of noncoding sequences between mouse and human. A new **alignment algorithm** was developed to cope with the fact that large parts of noncoding sequences are not **alignable** in a meaningful way because of genetic drift. This new **algorithm**, DNA Block **Aligner** (DBA), finds colinear-conserved blocks that are flanked by nonconserved sequences of varying lengths. The noncoding regions of the data set were **aligned** with DBA. The proportion of the noncoding regions covered by blocks >60% identical was 36% for upstream regions, 50% for 5' UTRs, 23% for introns, and 56% for 3' UTRs. These blocks of high identity were more or less evenly distributed across the length of the features, except for upstream regions in which the first 100 bp upstream of the transcription start site was covered in up to 70% of the gene pairs. This data set complements earlier sets on the basis of **cDNA** sequences and will be useful for further comparative studies.

RE.CNT 29 THERE ARE 29 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 7 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1999:374547 CAPLUS Full-text
DN 131:180381
TI Calign: **aligning** sequences with restricted affine gap penalties
AU Chao, Kun-Mao
CS Department of Computer Science and Information Management, Providence
University, Taichung, Taiwan
SO Bioinformatics (1999), 15(4), 298-304
CODEN: BOINFP; ISSN: 1367-4803
PB Oxford University Press
DT Journal
LA English
AB Given a **genomic** DNA sequence, it is still an open problem to determine its coding regions, i.e. the region consisting of exons and introns. The comparison of **cDNA** and **genomic**

DNA helps the understanding of coding regions. For such an application, it might be adequate to use the restricted affine gap penalties which penalize long gaps with a constant penalty. Several techniques developed for solving the approx. string-matching problem are employed to yield efficient **algorithms** for computing the optimal **alignment** with restricted affine gap penalties. In particular, efficient **algorithms** can be derived based on the suffix automaton with failure transitions and on the diagonal-wise monotonicity of the cost tables. We have implemented the above methods in C on Sun workstations running SunOS Unix. Preliminary expts. show that these approaches are very promising for **aligning** a **cDNA** sequence with a **genomic** DNA sequence.

RE.CNT 36 THERE ARE 36 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 8 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:785239 CAPLUS Full-text
DN 130:149220
TI A strategy for finding regions of similarity in complete genome sequences
AU Vincens, Pierre; Buffat, Laurent; Andre, Cecile; Chevrolat, Jean-Paul;
Boisvieux, Jean-Francois; Hazout, Serge
CS Departement de Biologie (FR 36), Ecole Normale Superieure, Paris, 75230,
Fr.
SO Bioinformatics (1998), 14(8), 715-725
CODEN: BOINFP; ISSN: 1367-4803
PB Oxford University Press
DT Journal
LA English
AB Complete **genomic** sequences will become available in the future. New methods to deal with very large sequences (sizes beyond 100 kb) efficiently are required. One of the main aims of such work is to increase our understanding of genome organization and evolution. This requires studies of the locations of regions of similarity. We present here a new tool, ASSIRC ('Accelerated Search for SImilarity Regions in Chromosomes'), for finding regions of similarity in **genomic** sequences. The method involves three steps: (i) identification of short exact chains of fixed size, called 'seeds', common to both sequences, using hashing functions; (ii) extension of these seeds into putative regions of similarity by a 'random walk' procedure; (iii) final selection of regions of similarity by assessing **alignments** of the putative sequences. We used simulations to **estimate** the proportion of regions of similarity not detected for particular region sizes, base identity proportions and seed sizes. This approach can be tailored to the user's specifications. We looked for regions of similarity between two yeast chromosomes (V and IX). The efficiency of the approach was compared to those of conventional programs BLAST and FASTA, by assessing CPU time required and the regions of similarity found for the same data set.

RE.CNT 25 THERE ARE 25 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 9 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:236053 CAPLUS Full-text
DN 129:13088
TI Data transferability from model organisms to human beings: insights from
the functional genomics of the flightless region of Drosophila
AU Maleszka, R.; De Couet, H. G.; Miklos, George L. Gabor
CS Research School of Biological Sciences, Australian National University,
Canberra, ACT 2600, Australia
SO Proceedings of the National Academy of Sciences of the United States of
America (1998), 95(7), 3731-3736
CODEN: PNASA6; ISSN: 0027-8424
PB National Academy of Sciences
DT Journal
LA English
AB At what biol. levels are data from single-celled organisms akin to a Rosetta stone for multicellular ones. To examine this question, we characterized a saturation-mutagenized 67-kb region of the Drosophila genome by gene deletions, transgenic rescues, phenotypic dissections, **genomic** and **cDNA** sequencing, bio-informatic anal., reverse transcription-PCR studies, and evolutionary comparisons. Data anal. using **cDNA/genomic** DNA **alignments** and bio-informatic **algorithms** revealed 12 different predicted proteins, most of which are absent from bacterial databases, half of which are absent from Saccharomyces cerevisiae, and nearly all of which have relatives in Caenorhabditis elegans and Homo sapiens. Gene order is not evolutionarily conserved; the closest relatives of these genes are scattered throughout the yeast, nematode, and human genomes. Most gene expression is pleiotropic, and deletion studies reveal that a morphol. phenotype is seldom observed when these genes are removed from the genome. These data pinpoint some general bottle-necks in functional genomics, and they reveal the acute emerging difficulties with data transferability above the levels of genes and proteins, especially with complex human phenotypes. At these higher levels the Rosetta stone analogy has almost no applicability. However, newer transgenic technologies in Drosophila and Mus, combined with coherency pattern analyses of gene networks, and synthetic neural modeling, offer insights into organismal function. We conclude that industrially scaled robogenomics in model organisms will have great impact if it can be realistically linked to epigenetic analyses of human variation and to phenotypic analyses of human diseases in different genetic backgrounds.

RE.CNT 39 THERE ARE 39 CITED REFERENCES AVAILABLE FOR THIS RECORD
ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 10 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1998:231829 CAPLUS Full-text
DN 129:36862
TI EbEST: an automated tool using expressed sequence tags to delineate gene
structure
AU Jiang, Jian; Jacob, Howard J.
CS Dep. Physiol., Lab. Genetics Res., Med. Coll. Wisconsin, Milwaukee, WI,
53226, USA
SO Genome Research (1998), 8(3), 268-275
CODEN: GEREFS; ISSN: 1088-9051
PB Cold Spring Harbor Laboratory Press
DT Journal
LA English
AB Large nos. of expressed sequence tags (ESTs) continue to fill public and private databases with partial **cDNA** sequences. However, using this huge amount of ESTs to facilitate gene finding in **genomic** sequence imposes a challenge, especially to wet-lab scientists who often have limited computing resources. In an effort to consolidate the information hidden in the vast number of ESTs into a readable and manageable format, we have developed EbEST - a program that automates the process of using ESTs to help delineate gene structure in long stretches of

genomic sequence. The EbEST program consists of three functional modules - the first module separates homologous ESTs into clusters and identifies the most informative ESTs within each cluster; the second module uses the informative ESTs to perform gapped **alignment** and to predict the exon-intron boundary; and the third module generates text file and graphic outputs that illustrate the orientation, exonic structure, and untranslated regions (UTRs) of putative genes in the **genomic** sequence being analyzed. Evaluation of EbEST with 176 human genes from the ALLSEQ set indicated that it performed in-line with several existing gene finding programs, but was more tolerant to sequencing errors. Furthermore, when EbEST was challenged with query sequences that harbor more than one gene, it suffered only a slight drop in performance, whereas the performance of the other programs evaluated decreased more. EbEST may be used as a stand-alone tool to annotate human **genomic** sequences with **EST**-derived gene elements, or can be used in conjunction with computational gene-recognition programs to increase the accuracy of gene prediction.

RE.CNT 16 THERE ARE 16 CITED REFERENCES AVAILABLE FOR THIS RECORD

ALL CITATIONS AVAILABLE IN THE RE FORMAT

L7 ANSWER 11 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1997:607203 CAPLUS Full-text
DN 127:288703
TI **EST**-GENOME: a program to **align** spliced DNA sequences to unspliced **genomic** DNA
AU Mott, Richard
CS Informatics Group, Sanger Cent., Welcome Trust Genome Campus, Cambridge, CB10 1SA, UK
SO CABIOS, Computer Applications in the Biosciences (1997), 13(4), 477-478
CODEN: COABER; ISSN: 0266-7061
PB Oxford University Press
DT Journal
LA English
AB A computer program **EST**-GENOME was developed for **aligning** spliced DNA to unspliced **genomic** DNA.

L7 ANSWER 12 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:317355 CAPLUS Full-text
DN 125:3768
TI Thermodynamic prediction of conserved secondary structure: application to the RRE element of HIV, the tRNA-like element of CMV and the **mRNA** of prion protein
AU Lueck, Rupert; Steger, Gerhard; Riesner, Detlev
CS Institut Physikalische Biologie, Heinrich-Heine-Universitaet, Duesseldorf, Germany
SO Journal of Molecular Biology (1996), 258(5), 813-826
CODEN: JMOBAK; ISSN: 0022-2836
PB Academic
DT Journal
LA English
AB An **algorithm** for prediction of conserved secondary structure of single-stranded RNA is presented. For each RNA of a set of homologous RNAs optimal and suboptimal secondary structures are calculated and stored in a base-pair probability matrix. A multiple sequence **alignment** is performed for the set of RNAs. The resulting gaps are introduced into the individual probability matrixes. These homologous probability matrixes are summed to give a consensus probability matrix emphasizing the conserved secondary structure elements of the RNA set. Thus the **algorithm** combines the advantages of thermodn. structure prediction by energy minimization with the information obtained from phylogenetic **alignment** of sequences. The **algorithm** is applied to three examples. The REV-responsive element of HIV, the structure of which is well known from the literature, was chosen to test the **algorithm.** The second example is the 3' terminal segment of **genomic** single-stranded RNAs of cucumber mosaic viruses; a structure similar to that of the related brome mosaic virus was expected and was confirmed. The third example is the prion-protein **mRNA** from different organisms; the structure of this **mRNA** is not known. By application of the **algorithm** highly conserved hairpins were found in the prion-protein **mRNA.**

L7 ANSWER 13 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:206781 CAPLUS Full-text
DN 124:311694
TI A Bayesian evolutionary distance for parametrically **aligned** sequences
AU Agarwal, Pankaj; States, David J.
CS Inst. Biomed. Comput., Washington Univ., St. Louis, MO, 63110, USA
SO Journal of Computational Biology (1996), 3(1), 1-17
CODEN: JCOBEM; ISSN: 1066-5277
PB Liebert
DT Journal
LA English
AB There is an inherent relation between the process of pairwise sequence **alignment** and the estimation of evolutionary distance. This relation is explored and made explicit. Assuming an evolutionary model and given a specific pattern of observed base mismatches, the relative probabilities of evolution at each evolutionary distance are computed using a Bayesian framework. The mean or the median of this probability distribution provides a robust **estimate** of the central value. The evolutionary distance has traditionally been computed as zero for an observed homol. of 20 bases with no mismatches; the authors prove that it is highly probable that the distance is >0.01. The mean of the distribution is 0.047, which is a better **estimate** of the evolutionary distance. Bayesian ests. of the evolutionary distance incorporate arbitrary prior information about variable mutation rates both over time and along sequence position, thus requiring only a weak form of the mol.-clock hypothesis. The endpoints of the similarity between **genomic** DNA sequences are often ambiguous. The probability of evolution at each evolutionary distance can be estimated over the entire set of **alignments** by choosing the best **alignment** at each distance and the corresponding probability of duplication at that evolutionary distance. A central value of this distribution provides a robust evolutionary distance **estimate** The authors provide an efficient **algorithm** for computing the parametric **alignment**, considering evolutionary distance as the only parameter. These techniques and ests. are used to infer the duplication history of the **genomic** sequence in C. elegans and in S. cerevisiae. The results indicate that repeats discovered using a single scoring matrix show a considerable bias in subsequent evolutionary distance ests.

L7 ANSWER 14 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1996:168439 CAPLUS Full-text
DN 124:221863
TI Improved tools for DNA comparison and clustering
AU Parsons, J.D.

CS School of Medicine, Washington University, St Louis, MO, 63108, USA
SO Computer Applications in the Biosciences (1995), 11(6), 603-13
CODEN: COABER; ISSN: 0266-7061
PB Oxford University Press
DT Journal
LA English
AB DNA sequence clustering is an effective aid of the comprehension, summarization and compression of DNA sequence databases. Previous work created programs suitable for the comparison and clustering of **cDNA** sequences but new enhanced programs have been written to cluster **genomic** DNA fragments, large **EST** projects, and entire DNA databases. Three new programs (ICAtools) are discussed: ICAass, N2tool, and ICAmatches. ICAass has been used to compress the EMBL database by hiding or removing sequences with various degrees of redundancy. It also has the fastest database querying mode. N2tool provides fast and sensitive clustering of **genomic** fragment databases on the basis of small areas of local similarity. N2tool has proven utility in the discovery of contaminating vector or other artifactual sequence when the potential contaminant is not otherwise known. ICAmatches is a new cluster anal. program that uses a novel **alignment** style to present multiple **alignment** summaries. All the tools are convenient to use because they share a common memoryfrugal index format and accept most DNA sequence formats directly.

L7 ANSWER 15 OF 15 CAPLUS COPYRIGHT 2004 ACS on STN
AN 1993:663583 CAPLUS Full-text
DN 119:263583
TI A quality control **algorithm** for DNA sequencing projects
AU White, Owen; Dunning, Ted; Sutton, Granger; Adams, Mark; Venter, J. Craig;
Fields, Chris
CS Inst. Genomic Res., Gaithersburg, MD, 20878, USA
SO Nucleic Acids Research (1993), 21(16), 3829-38
CODEN: NARHAD; ISSN: 0305-1048
DT Journal
LA English
AB Heterologous DNA sequences from rearrangements with the genomes of host cells, **genomic** fragments from hybrid cells, or impure tissue sources can threaten the purity of libraries that are derived from RNA or DNA. Hybridization methods can only detect contaminants from known or suspected heterologous sources, and whole library screening is tech. very difficult. Detection of contaminating heterologous clones by sequence **alignment** is only possible when related sequences are present in a known database. The authors have developed a statistical test to identify heterologous sequences that is based on the differences in hexamer composition of DNA from different organisms. This test does not require that sequences similar to potential heterologous contaminants are present in the database, and can in principle detect contamination by previously unknown organisms. The authors have applied this test to the major public expressed sequence tag (**EST**) data sets to evaluate its utility as a quality control measure and a peer evaluation tool. There is detectable heterogeneity in most human and C. elegans **EST** data sets but it is not apparently associated with cross-species contamination. However, there is direct evidence for both yeast and bacterial sequence contamination in some public database sequences annotated as human. Results obtained with the hexamer test have been confirmed with similarity searches using sequences from the relevant data sets.

=> log y

COST IN U.S. DOLLARS	SINCE FILE ENTRY	SESSION	TOTAL
FULL ESTIMATED COST		72.56	72.77

DISCOUNT AMOUNTS (FOR QUALIFYING ACCOUNTS)	SINCE FILE ENTRY	SESSION	TOTAL
CA SUBSCRIBER PRICE		-10.40	-10.40

STN INTERNATIONAL LOGOFF AT 19:13:01 ON 12 MAY 2004